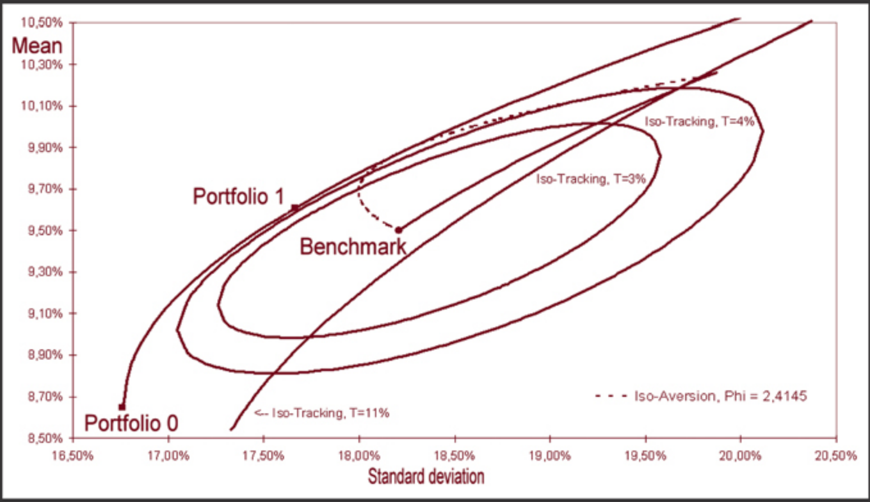
Jean-Luc Prigent

Portfolio Optimization and Performance Analysis



Chapman & Hall/CRC FINANCIAL MATHEMATICS SERIES

CHAPMAN & HALL/CRC FINANCIAL MATHEMATICS SERIES

Portfolio Optimization and Performance Analysis

CHAPMAN & HALL/CRC
Financial Mathematics Series

Aims and scope:
The field of financial mathematics forms an ever-expanding slice of the financial sector. This series aims to capture new developments and summarize what is known over the whole spectrum of this field. It will include a broad range of textbooks, reference works and handbooks that are meant to appeal to both academics and practitioners. The inclusion of numerical code and concrete real-world examples is highly encouraged.

Published Titles

American-Style Derivatives; Valuation and Computation, *Jerome Detemple*

Financial Modelling with Jump Processes, *Rama Cont and Peter Tankov*

An Introduction to Credit Risk Modeling, *Christian Bluhm, Ludger Overbeck, and Christoph Wagner*

Portfolio Optimization and Performance Analysis, *Jean-Luc Prigent*

Robust Libor Modelling and Pricing of Derivative Products, *John Schoenmakers*

Structured Credit Portfolio Analysis, Baskets & CDOs, *Christian Bluhm and Ludger Overbeck*

Proposals for the series should be submitted to one of the series editors above or directly to:
CRC Press, Taylor and Francis Group
24-25 Blades Court
Deodar Road
London SW15 2NU
UK

CHAPMAN & HALL/CRC FINANCIAL MATHEMATICS SERIES

Portfolio Optimization and Performance Analysis

Jean-Luc Prigent

Chapman & Hall/CRC
Taylor & Francis Group
Boca Raton London New York

Chapman & Hall/CRC is an imprint of the
Taylor & Francis Group, an **informa** business

Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 1-58488-578-5 (Hardcover)
International Standard Book Number-13: 978-1-58488-578-8 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Prigent, Jean-Luc, 1958-
Portfolio optimization and performance analysis / Jean-Luc Prigent.
p. cm. -- (Chapman & Hall/CRC financial mathematics series ; 7)
Includes bibliographical references and index.
ISBN-13: 978-1-58488-578-8 (alk. paper)
ISBN-10: 1-58488-578-5 (alk. paper)
1. Portfolio management. 2. Investment analysis. 3. Hedge funds. I. Title. II. Series.

HG4529.5.P735 2007
332.6--dc22 2006100727

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Preface

Since the seminal mean-variance analysis was introduced by Markowitz (1952), the portfolio management theory has been expanded to take account of different features:

- Dynamic portfolio optimization as per Merton (1962);
- Choice of new decision criteria, based on risk aversion (utility functions) or risk measures (VaR, CVaR and beyond);
- Market imperfections, *e.g.*, transaction costs; and,
- Specific portfolio strategies, such as portfolio insurance or alternative methods (hedge-funds).

At the same time, many new financial products has been introduced, based in particular on financial derivatives.

Due to this intensive development and increasing complexity, this book has four purposes:

- First, to recall standard results and to provide new insights about the axiomatics of the individual choice in an uncertain framework. A concise introduction to portfolio choice under uncertainty based on investors' preferences (usually represented by utility functions), and on several kinds of risk measures. These theories are the fundamental basis of portfolio optimization.

  - Chapter 1 recalls the seminal approach of the utility maximization, introduced by Von Neumann and Morgenstern. It also deals with further extensions of this theory, such as weighted expected utility theory, non-expected utility theory, *etc.*

  - Chapter 2 contains a survey about a new approach: the risk measure minimization. Such risk measures have been recently introduced in particular to take better account of nonsymmetric asset return distributions.

- Second, to provide a precise overview on standard portfolio optimization. Both passive and active portfolio management are considered. Other results, such as risk measure minimization, are more recent.

- Chapter 3 is devoted to the very well-known Markowitz analysis. Some extensions are analyzed, in particular with risk minimization constraints such as safety criteria.

- Chapter 4 deals with two important standard fund managements: managing indexed funds and benchmarked portfolio optimization. In particular, statistical methods to replicate a financial index are detailed and discussed. As regards benchmarking, the tracking error is computed and analyzed.

- Chapter 5 recalls results about the main performance measures, such as the Sharpe and Treynor ratios and the Jensen alpha.

• Third, to make accessible the literature about stochastic optimization applied to mathematical finance (see for example Part III) to students, to researchers who are not specialists on this subject, and to financial engineers. In particular, a review of the main standard results both for static and dynamic cases are provided. For this purpose, precise mathematical statements are detailed without "too many" technicalities. In particular:

- Chapter 6 provides an introduction to dynamic portfolio optimization. The two main methods are the theory of stochastic control based on dynamic programming principle and, more recently, the martingale approach jointly used with convex duality.

- Chapter 7 gives two important applications of previous results: the search for an optimal portfolio profile and the long-term management.

- Chapter 8 is the more "technical" one. It provides an overview on portfolio optimization with market frictions, such as incompleteness, transaction costs, labor income, random time horizon, *etc.*

• Finally, to show how theoretical results can be applied to practical and operational portfolio optimization (Part IV). This last part of the book deals with structured portfolio management which has grown significantly in the past few years.

- Chapter 9 is devoted to portfolio insurance and, in particular, to OBPI and CPPI strategies.

- Chapter 10 shows how common strategies, used by practitioners, may be justified by utility maximization under, for example, guarantee constraints. It summarizes the main results concerning optimal portfolios when risk measures such as expected shortfall are introduced to limit downside risk.

- Chapter 11 recalls some problems when dealing with hedge funds, in particular the choice of appropriate performance measures.

As a by-product, special emphasis is put on:

- Utility theory versus practice;
- Active versus passive management; and,
- Static versus dynamic portfolio management.

I hope this book will contribute to a better understanding of the modern portfolio theory, both for students and researchers in quantitative finance.

I am grateful to the CRC editorial staff for encouraging this project, in particular Sunil Nair, and for the help during the preparation of the final version: Michele Dimont and Shashi Kumar.

Jean-Luc PRIGENT, PARIS, February 2007.

Contents

List of Tables

List of Figures

Part I

Utility and risk analysis

"[Under uncertainty] there is no scientific basis on which to form any calculable probability whatever. We simply do not know. Nevertheless, the necessity for action and for decision compels us as practical men to do our best to overlook this awkward fact and to behave exactly as we should if we had behind us a good Benthamite calculation of a series of prospective advantages and disadvantages, each multiplied by its appropriate probability waiting to be summed."

John Maynard Keynes, "General Theory of Employment," *Quarterly Journal of Economics*, (1937).

Nowadays, financial theory is one of the major economic fields where decision-making under uncertainty plays a crucial part. Actually, many sources of risk (market, model, liquidity, operational, *etc.*) have to be taken into account and carefully examined for most financial activities, such as pricing and hedging derivatives, asset allocation, or credit portfolio management.

Assume that these risky events are identified with, for example, probability distributions that may be objective or subjective. Nevertheless:

How can we model individual decisions under uncertainty?

Is it possible to rationalize traders or portfolio managers strategies? Can we provide them with sufficiently operational and computational tools to try to improve their decision process?

As it is well-known, a unified framework can be proposed to quantify uncertainty in financial modelling: the utility theory, and especially the expected utility theory introduced by John von Neumann and Oskar Morgenstern in [400] and recognized for its usefulness and applicability.

Utility functions are based on risk aversion modelling from which the notion of risk premium can be defined. In Chapter 1, basic notions of the theory of decision under uncertainty are recalled. The emphasis is put on the expected utility theory and various risk aversion notions. One of the advantages of the expected utility is that it provides an operational tool to determine explicit portfolios under mild assumptions. In this framework, the risk-aversion allows a calibration of the portfolio weights, as detailed in Part III.

Nevertheless, the increasing development of the so-called *behavorial economics and finance*, based on empirical evidence, justifies sections devoted to alternative preference representation theories. Indeed, many experimental studies have shown that individuals (in particular the investors) do not act according to the expected utility theory. This can partly explain investment anomalies such as insufficient diversification, financial bubbles, *etc.*

However a new stream has emerged based on bank activity regulation. It focuses in particular on potential losses and downside risk.

In [373] and [374], Markowitz proposed to measure risk of portfolio returns by means of their variances which involve judiciously the joint distribution of returns of all assets. Despite its simplicity and tractability, the Markowitz model has two pitfalls:

- First, the probability distribution of each asset return is characterized only by its first two moments. In the case of nonGaussian distributions

(even symmetrical), the Markowitz model and utility theories are mainly compatible for quadratic utility functions.

- Second, the dependence structure is only described by the linear correlation coefficients of each pair of asset returns. As shown, for example, by Alexander [14], the linear correlation coefficient is not always applicable. It also may imply incorrect results when probability distributions are not elliptic (see Joe [307]), as proved for instance by Embrechts et al. ([201] and [202]). In that case, severe losses can be observed if extreme events are too underestimated.

What kind of risk measures can be introduced?

Unlike *dispersion* risk measures such as the standard deviation, other measures have been proposed, based rather on *downside risks.*

From the seminal paper by Artzner, Delbaen, Eber and Heath [31], specific axioms have been introduced to model risk measures (*coherent*), and further examined and generalized by Föllmer and Schied [236] (*convex* measures).

Chapter 2 is devoted to the definitions and main properties of such risk measures. Note that, as for preference representation, the theory of risk measures is not yet achieved, in particular when they have to be defined in a dynamic framework. Besides, both approaches are linked, as shown by recent results. Among the possible operational risk measures, the value-at-risk and its "coherent" extension, the expected shortfall, have emerged as important tools to bank regulation and risk management.

This is the reason why in Chapter (2), some emphasis is put on these measures, in particular on some results about their estimation and sensitivities computation. Under some additional and "rational" specific axioms, risk measures can be defined from the expected shortfall (the so-called *spectral risk measures*).

Portfolio management can also involve such measures to limit risk exposure, as detailed for instance in Part IV, Chapter 10.

Chapter 1

Utility theory

The importance of risk and uncertainty in economic analysis was suggested for the first time by Frank H. Knight in his seminal treatise *Risk, Uncertainty and Profit* (see [328]). Previously, very few economists considered that risk and uncertainty might play a key role in economic theory, except for some notable examples like Carl Menger [384], Irving Fisher [227] and Francis Edgeworth [186]. The problem was:

- First, to define precisely the notions of "uncertainty" or "risk" when events are random; and,
- Second, to model the choice process within risk and uncertainty.

The notion of choice under risk and uncertainty was not well modelled for a long time, despite the results of Hicks [293] and Marschak [377], who understood that preferences should be defined also on distributions, to take account simultaneously of the evaluation of the level of risk or uncertainty, and of the pure preferences over outcomes. However, Bernoulli [54] had formerly introduced the notion of *expected utility* to solve the famous St.Petersburg paradox posed in 1713. The expected utility theory allows the representation of the level of satisfaction by the sum of utilities from outcomes weighted by the probabilities of these outcomes. Nevertheless, in that case, a gain can increase utility less than a decline can reduce it, which was not considered as rational.

In the seminal *Theory of Games and Economic Behavior*, John von Neumann and Oskar Morgenstern [400] succeeded in providing a rational foundation for decision-making under risk according to expected utility properties. At this time, this axiomatic foundation for decision-making under risk was not well understood. This theory was further developed by Marschak [378], Samuelson ([445] and [446]), Herstein and Milnor [291] and others.

The expected utility hypothesis was rehanced in the famous *Foundations of Statistics* by Savage [449]. Savage proposed to deduce the expected utility property without imposing prior objective probabilities but by determinating implicit subjective probabilities. This approach was further studied by Anscombe and Aumann [26]. Thus the von Neumann-Morgenstern theory was extended by the Savage-Anscombe-Aumann "subjective" approach.

The "state-preference" approach to uncertainty was introduced by Arrow [28] and Debreu [152]. It does not necessarily assume the existence of objective or subjective probabilities. Rather, it concerns actual goods than money amounts and has been applied to study general economic equilibria.

The notion of "risk aversion" was introduced by Friedman and Savage [242] and by Markowitz [374]. The measures of risk aversion were examined by Pratt [412] and Arrow [29], and later studied by Ross [434]. The different notions of risk aversion have been further developed by Yaari [506] and K-ihlstrom and Mirman [327]. The notions and definitions of "riskiness" based on stochastic dominance were suggested by Rothschild and Stiglitz ([435] and [436]) and Diamond and Stiglitz [167].

The expected utility assumption for modelling choice under risk and uncertainty has been discussed and disputed, in particular by Allais [19] and Ellsberg [196]. This debate has generated many alternative approaches to expected utility theory. Some of them have been based on experimental experiences, such as in Kahneman and Tversky [314]. The main alternative models are: *weighted expected utility* (Allais [20], Chew and McCrimmon [120]); *rank-dependent expected utility* (Quiggin [420], Yaari ([506], [507]), and more recently, *the cumulative prospect theory* by Kahneman and Tversky [496]); *non-linear expected utility* (Machina [367]); and *regret theory* (Loomes and Sugden [365]) to take account of preference reversals. Other alternative expected utility theories have been developed within the framework of Savage's subjective probability: *non-additive expected utility* (Schmeidler [454]) and *state-dependent preferences* (Karni [323]).

To summarize, the purpose of the decision theory is to provide analytical tools of different degrees of formality in order to model the behavior of a *decison maker* who has to choose among a set of *alternatives* with different *consequences*. Typically, since Knight [328], three cases are distinguished, according to the degree of information:

- First, the environment is *certain*: the agent perfectly knows the event that will occur in the future.
- Second, the environmment is *risky*: this means existence of uncontrollable random events for which the modelling of a probability space can be proposed, in particular a probability distribution can be determined.
- Finally, the environmment is *uncertain*: in that case, the probability distribution is unknown.

Financial theory is mainly concerned with the second situation. However, for the third case, note that under some assumptions a subjective probability may exist, as proved by Savage [449].

1.1 Preferences under uncertainty

In order to model any decision problem under risk, it is necessary to introduce a functional representation of preferences which measures the degree of satisfaction of the decision maker. This is the purpose of the utility theory. The investor is supposed to be "rational": this means that his choices are made according to given "good" rules which are "stable" over time (in some sense). Thus a binary relation on possible outcomes can be proposed to analyze his behavior. Specific axioms are introduced to describe his "rationality." Then, for this given identified choice functional, his optimal decision (for example his investment strategy) is determined from the "maximization" of this criterion.

1.1.1 Lotteries

Within a risky framework, first we must pick out all possible outcomes which may have an impact on the consequences of the decisions. Secondly, we must associate each random event with a probability.

Let Ω represent the set of possible outcomes. To simplify the exposition, we suppose Ω is finite:

$$\Omega = \{\omega_1, ..., \omega_m\}.$$

Let $p = \{p_1, ..., p_m\}$ be the probability of occurence of Ω:

$$\forall i, 0 \leq p_i \leq 1 \text{ and } \sum_i p_i = 1.$$

DEFINITION 1.1 *A lottery L is defined by a vector $\{(\omega_1, p_1), ..., (\omega_m, p_m)\}$. The set of all lotteries with the same given set of outcomes Ω is denoted by $\mathcal{L}$. A compound lottery L_c is a lottery whose outcomes are also lotteries.*

Consider for example a compound lottery with two outcomes: the lottery L^a and the lottery L^b with respective probabilities α and $1-\alpha$. Then the probability that the outcome of L_c will be ω_i is given by:

$$p_i = \alpha p_i^a + (1-\alpha)p_i^b.$$

Therefore, L_c has the same vector of probabilities as the convex combination:

$$\alpha L^a + (1-\alpha)L^b.$$

1.1.2 Axioms on preferences

The decision maker is assumed to be "rational" if his preference relation (denoted by $\succeq$) over the set of lotteries $\mathcal{L}$ is a binary relation which satisfies the following axioms:

Axiom 1 :

- The relation $\succeq$ is *complete* (all lotteries are always comparable by $\succeq$):

$$\forall L^a \in \mathcal{L},\, \forall L^b \in \mathcal{L},\; L^a \succeq L^b \text{ or } L^b \succeq L^a.$$

  The *indifference* relation $\sim$ associated to the relation $\succeq$ is defined by:

$$\forall L^a \in \mathcal{L},\, \forall L^b \in \mathcal{L},\; L^a \sim L^b \Longleftrightarrow L^a \succeq L^b \text{ and } L^b \succeq L^a.$$

- The relation $\succeq$ is *reflexive*:

$$\forall L \in \mathcal{L}, L \succeq L.$$

- The relation $\succeq$ is *transitive*:

$$\forall L^a \in \mathcal{L},\, \forall L^b \in \mathcal{L},\, \forall L^c \in \mathcal{L},\; L^a \succeq L^b \text{ and } L^b \succeq L^c \Longrightarrow L^a \succeq L^c.$$

Another standard assumption is the *continuity*: small changes on probabilities do not modify the ordering between two lotteries. This property is specified in the following axiom:

Axiom 2 : The preference relation $\succeq$ on the set $\mathcal{L}$ of lotteries is such that $\forall L^a \in \mathcal{L}$, $\forall L^b \in \mathcal{L}$, $\forall L^c \in \mathcal{L}$, if $L^a \succeq L^b \succeq L^c$ then there exists a scalar $\alpha \in [0,1]$ such that:

$$L^b \sim \alpha L^a + (1-\alpha)L^c.$$

This continuity axiom implies the existence of a functional $\mathcal{U} : \mathcal{L} \longrightarrow \mathbb{R}$ such that:

$$\forall L^a \in \mathcal{L},\, \forall L^b \in \mathcal{L},\; L^a \succeq L^b \Longleftrightarrow \mathcal{U}(L^a) \geq \mathcal{U}(L^b).$$

The previous axioms were already well-known in the economic theory of consumer choice (sometimes called the "weak order" axioms). To develop the analysis of economics under uncertainty, more properties must be imposed on the preferences. One of the most important conditions that can be added to describe the behavior of the economic agent (here the "investor") is the *independence axiom*, which is the foundation of the standard theory under uncertainty, but considered more troublesome (stated as in Jensen [301]):

Axiom 3 : The preference relation $\succeq$ on the set $\mathcal{L}$ of lotteries is such that $\forall L^a \in \mathcal{L}$, $\forall L^b \in \mathcal{L}$, $\forall L^c \in \mathcal{L}$ and for all $\alpha \in [0,1]$,

$$L^a \succeq L^b \Longleftrightarrow \alpha L^a + (1-\alpha)L^c \succeq \alpha L^b + (1-\alpha)L^c.$$

This property means that if two lotteries L^a and L^b are mixed in the same way with any third lottery L^c then the preference ordering between the two new mixed lotteries is not modified.

1.2 Expected utility

The independence axiom, implicitly introduced in von Neumann and Morgenstern [400], characterizes the expected utility criterion: indeed, it implies that the preference functional $\mathcal{U}$ on the lotteries must be linear in the probabilities of the possible outcomes:

THEOREM 1.1
Assume that the preference relation $\succeq$ on the set $\mathcal{L}$ of lotteries satisfies the continuity and independence axioms. Then, the relation $\succeq$ can be represented by a preference functional that is linear in probabilities: there exists a function u defined (up to a positive linear transformation) on the space of possible outcomes Ω and with values in $\mathbb{R}$ such that for any two lotteries $L^a = \{(p_1^a, ..., p_n^a)\}$ and $L^b = \{(p_1^b, ..., p_n^b)\}$, we have:

$$L^a \succeq L^b \Longleftrightarrow \sum_{i=1}^{n} u(\omega_i) p_i^a \geq \sum_{i=1}^{n} u(\omega_i) p_i^b. \tag{1.1}$$

PROOF In what follows, only a sketch of the proof is presented. We have mainly to prove that for any two lotteries L^a and L^b, and any compound lottery $L = \alpha L^a + (1-\alpha) L^b$:

$$\mathcal{U}[\alpha L^a + (1-\alpha) L^b] = \alpha \mathcal{U}[L^a] + (1-\alpha) \mathcal{U}[L^b].$$

- First, consider the worst and best lotteries, L^{wo} and L^{be} in $\mathcal{L}$, with respect to the preference functional $\mathcal{U}$. They are obtained by minimizing and maximizing $\mathcal{U}$ on a closed preference interval of $\mathcal{L}$. Then, for any lottery L in $\mathcal{L}$, we have : $L^{wo} \succeq L \succeq L^{be}$. Thus, by the continuity axiom, there exist two scalars β^a and β^b in $[0,1]$ such that:

  $$L^a \sim \beta^a L^{be} + (1-\beta^a) L^{wo} \text{ and } L^b \sim \beta^b L^{be} + (1-\beta^b) L^{wo}.$$

  Note that β^a and β^b are unique.

- (*Mixture monotonicity.*) We have:

  $$L^a \succeq L^b \Longleftrightarrow \beta^a \geq \beta^b.$$

  Indeed, if $\beta^a \geq \beta^b$, then the parameter $\beta = \frac{\beta^a - \beta^b}{1-\beta^b}$ is in $[0,1]$. Then, we have:

  $$\beta^a L^{wo} + (1-\beta^a) L^{be} \sim \beta L^{be} + (1-\beta)[\beta^b L^{be} + (1-\beta^b) L^{wo}].$$

Moreover, by definition of L^{be}, $L^{be} \succeq \beta^b L^{be} + (1-\beta^b)L^{wo}$. Therefore, using the independence axiom, we deduce:

$$\beta L^{be} + (1-\beta)[\beta^b L^{be} + (1-\beta^b)L^{wo}] \succeq$$
$$\beta[\beta^b L^{be} + (1-\beta^b)L^{wo}] + (1-\beta)[\beta^b L^{be} + (1-\beta^b)L^{wo}].$$

Consequently, $L^a = \beta^a L^{be} + (1-\beta^a)L^{wo} \succeq L^b = \beta^b L^{be} + (1-\beta^b)L^{wo}$. This result is quite intuitive, since it means that if we construct two compound lotteries L^a and L^b with different weights, then we prefer the compound lottery in which the best lottery is given the higher weight.

- Consequently, we conclude that the functional $\mathcal{U}$, such that for any lottery $L \sim \beta L^{be} + (1-\beta)L^{wo}$ we have $\mathcal{U}(L) = \beta$, satisfies exactly Definition (1.1.2) of a preference functional associated to the preference relation $\succeq$.

- Finally, we must prove that:

$$\mathcal{U}[\alpha L^a + (1-\alpha)L^b] = \alpha\beta^a + (1-\alpha)\beta^b.$$

This is equivalent to show that:

$$\alpha L^a + (1-\alpha)L^b \sim (\alpha\beta^a + (1-\alpha)\beta^b)L^{be} + (\alpha(1-\beta^a) + (1-\alpha)(1-\beta^b))L^{wo}.$$

Using twice the independent axiom, we get:

$$\alpha L^a + (1-\alpha)L^b \sim \alpha[\beta^a L^{be} + (1-\beta^a)L^{wo}] + (1-\alpha)L^b$$
$$\sim \alpha[\beta^a L^{be} + (1-\beta^a)L^{wo}] + (1-\alpha)[\beta^b L^{be} + (1-\beta^b)L^{wo}]$$
$$\sim (\alpha\beta^a + (1-\alpha)\beta^b)L^{be} + (\alpha(1-\beta^a) + (1-\alpha)(1-\beta^b))L^{wo}.$$

Then, the previous result is extended to the whole space of lotteries $\mathcal{L}$ and the proof is finished.

$\square$

REMARK 1.1

1) The property in Theorem (1.1) is an equivalence, i.e. expected utility implies the three axioms.

2) Note that the previous results have been proved for simple lotteries, i.e. probability distributions which take positive values only for a finite number of outcomes. We have to extend these results to continuous spaces ("infinite support"): i.e. for any probability measure $\mathbb{P}$ on Ω, we must prove an analogue to the expected utility decomposition

$$\mathcal{U}(\mathbb{P}) = \int_\Omega u(\omega)d\mathbb{P}(\omega). \quad (1.2)$$

For this purpose, the continuity axiom (or Archimedian axiom) has to be supplemented as shown in Fishburn [223]. $\square$

1.3 Risk aversion

When facing alternatives with comparable returns, what are the attitudes of investors towards risk? Despite, for example, state-owned lotteries, usual observations on financial or insurance markets show that generally human beings are risk-averse. For example, they choose to invest on risky assets only if their expected returns are significantly larger than the riskless one. To illustrate this notion, consider the construction of Friedman and Savage [242]: let X a random variable with only two values x_1 and x_2, and let p, the probability of x_1, and $(1-p)$ be the probability of x_2. Let u represent a utility function defined on the outcomes. Consider the following two lotteries L^a and L^b: lottery L^a pays the amount $\mathbb{E}[X]$ with probability equal to 1. Lottery L^b pays x_1 with probability p, and x_2 with probability $(1-p)$. These lotteries have the same expected income, but an investor who is averse to risk would select L^a instead of L^b. Therefore, we have:

$$\mathcal{U}[L^a] = u(\mathbb{E}[X]) \, and \, \mathcal{U}[L^b] = \mathbb{E}[u(X)]. \tag{1.3}$$

Then, as illustrated in Figure (1.1), the concavity of u implies that the utility of expected return $u(\mathbb{E}[X])$ is greater than the expected utility $\mathbb{E}[U(X)]$.

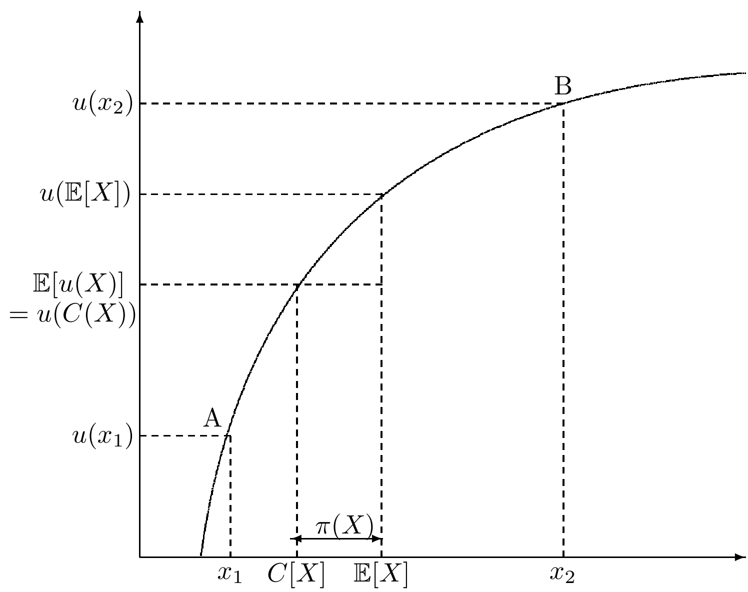


FIGURE 1.1: Risk aversion, certainty equivalence, and concavity

Indeed, by the definition of concavity, we have:

$$\forall \lambda \in [0,1], \forall a \in \mathbb{R}, \forall b \in \mathbb{R}, \lambda u(a) + (1-\lambda)u(b) \leq u(\lambda a + (1-\lambda)b).$$

The basic mathematical result is the well-known Jensen's inequality: consider a concave real-valued function f and a real-valued random variable Y with finite expectation. Then we have $\mathbb{E}[f(Y)] \leq f(\mathbb{E}[Y])$.

Another way to represent this risk-aversion is to introduce the *certainty equivalent* of the lottery L^b. This lottery, denoted by $C[X]$, is the sure or *certainty-equivalent* lottery which yields the same utility as the random lottery L^b. Thus, the investor is indifferent to the choice of receiving $C[X]$ with certainty or investing on the risky lottery L^b with expected return $\mathbb{E}[X]$. The risk-aversion is equivalent to the inequality $C[X] < \mathbb{E}[X]$. The difference $\pi[X] = \mathbb{E}[X] - C[X]$ is called the *risk-premium* as introduced in Pratt [412]. It is the maximum amount that the investor accepts to lose in order to get a riskless income. From these properties, we can propose the following definitions:

DEFINITION 1.2
1) An investor is risk-averse if $C[X] \leq \mathbb{E}[X]$, or equivalently if $\pi[X] \geq 0$, for all random variables X.
2) An investor is risk-neutral if $C[X] = \mathbb{E}[X]$, or equivalently if $\pi[X] = 0$, for all random variables X.
3) An investor is risk-loving if $C[X] \geq \mathbb{E}[X]$, or equivalently if $\pi[X] \leq 0$, for all random variables X.

Using the properties of concavity/convexity of utility functions, we deduce a characterization of the risk-aversion:

THEOREM 1.2
Let u be a utility function representing preferences over the set of outcomes. Assume that u is increasing. Then:

1) The function u is concave if and only if the investor is risk-averse.
2) The function u is linear if and only if the investor is risk-neutral.
3) The function u is convex if and only if the investor is risk-loving.

PROOF Consider for example the concavity case. Then: $\forall \lambda \in [0,1], \forall a \in \mathbb{R}, \forall b \in \mathbb{R}$, $\lambda u(a) + (1-\lambda)u(b) \leq u(\lambda a + (1-\lambda)b)$. But $\mathbb{E}[X] = \lambda a + (1-\lambda)b$ and $\mathbb{E}[u(X)] = \lambda u(a) + (1-\lambda)u(b)$. By definition, $u(C[X]) = \mathbb{E}[u(X)]$. Thus, $u(C[X]) \leq u(\mathbb{E}[X])$. Therefore, since u is increasing, we deduce: $C[X] \leq \mathbb{E}[X]$, which is the definition of risk-aversion.

$\square$

REMARK 1.2 As noted in [242], an investor's utility function may have different curvatures: For example, he may be risk-averse for small and very high wealth levels, but risk-loving for intermediate levels. In that case, his utility function has a double inflection. However, such behavior modelling can be criticized as was done by Markowitz [374]. □

1.3.1 Arrow-Pratt measures of risk aversion

Human beings preferences are heterogeneous: some may prefer safety to risk, others do not. In the first case, they will invest significantly on "riskless" assets, treasury bonds for example. In the second case, they will purchase stocks that will represent a high proportion of their financial portfolios. But, how can we measure the "degree" of risk aversion of an investor? Since for example utility functions are defined up to linear transformations, the concavity itself is not sufficient to characterize this degree. Another possible approach is to examine the risk premia and to relate them to concavity. This is the way chosen by Pratt [412] and Arrow [29]. Consider the following result due to Pratt:

DEFINITION 1.3 *Let u and v be two utility functions representing preferences over wealth. The preference u has more risk-aversion than v if the risk-premia satisfy: $\pi_u(X) \geq \pi_v(X)$, for all random real-valued variables X.*

THEOREM 1.3
Let u and v be two utility functions representing preferences over wealth. Assume that they are continuous, monotonically increasing, and twice differentiable. Then the following properties are equivalent and characterize the "more risk-aversion":

1) The derivatives of both utility functions are such that: $-\frac{u"}{u'}(x) > -\frac{v"}{v'}(x)$, for every x in $\mathbb{R}$.

2) There exists a concave function Φ such that: $u(x) = \Phi[v(x)]$, for every x in $\mathbb{R}$.

3) The risk-premia satisfy: $\pi_u(X) \geq \pi_v(X)$, for all random real-valued variables X.

PROOF Case 1: (1)⇒(2).
Since the function v is monotonic and continuous, v has an inverse v^{-1}. Define the function Φ by: $\Phi(y) = u \circ v^{-1}(y)$. Then, by construction, we have:

$$u(x) = \Phi[v(x)].$$

Since the functions u and v are twice differentiable, we deduce that Φ is also twice differentiable and:

$$u'(x) = \Phi'[v(x)]v'(x).$$

Thus, Φ' is non-negative. Differentiating again, we get:

$$u\text{''}(x) = \Phi\text{''}[v(x)]v'^2(x) + \Phi'[v(x)]v\text{''}(x).$$

Therefore:

$$u\text{''}(x) = \Phi\text{''}[v(x)]v'^2(x) + u'(x)v\text{''}(x)/v'(x),$$

and finally we get:

$$v\text{''}(x)/v'(x) - u\text{''}(x)/u'(x) = (-\Phi\text{''}[v(x)])(v'^2(x)/u'(x)),$$

with $(v'^2(x)/u'(x)) > 0$. Now, from assumption (1), we have:

$$\frac{v\text{''}}{v'}(x) - \frac{u\text{''}}{u'}(x) > 0.$$

Consequently, Φ" is negative and Φ is concave.

Case 2: (2)$\Rightarrow$(3).
By definition of $C[X]$, we have: $u(C_u[X]) = \mathbb{E}[u(X)]$. Therefore, since $u(x) = \Phi[v(x)]$, then $u(C_u[X]) = \mathbb{E}[\Phi[v(X)]]$. Now, since Φ is concave, we have:

$$\mathbb{E}[\Phi[v(X)]] \leq \Phi[\mathbb{E}[v(X)]],$$

which implies:

$$u(C_u[X]) \leq \Phi[v(C_v[X])].$$

Finally, since $u(.) = \Phi[v(.)]$ is increasing, we deduce:

$$C_u[X] \leq C_v[X],$$

which implies: $\pi_u(X) \geq \pi_v(X)$.

Case 3: (3)$\Rightarrow$(1). Equivalently, we can prove that not(1)$\Rightarrow$not(3). If "not(1)", then we have:

$$-\frac{u\text{''}}{u'}(a) \;<\; -\frac{v\text{''}}{v'}(xa),$$

for some a in $\mathbb{R}$. By continuity, there exists a neighborhood $\mathcal{V}(a)$ for which this inequality still holds for all x $\in \mathcal{V}(a)$. Consider a random variable X with values in $\mathcal{V}(a)$ and 0 otherwise. First, we can prove that the previous tranformation Φ is convex on the set $\mathcal{V}(a)$: In the proof (1)$\Rightarrow$(2), we have seen that

$$\frac{v\text{''}}{v'}(x) - \frac{u\text{''}}{u'}(x) = -k\Phi[v(x)].$$

But on $(\mathcal{V})(a)$, we have

$$\frac{v\text{''}}{v'}(x) - \frac{u\text{''}}{u'}(x) > 0,$$

which implies that Φ" > 0 on $(\mathcal{V})(a)$. Second, since Φ is convex, using the result (2)$\Rightarrow$(3), we deduce: $\pi_u(X) \leq \pi_v(X)$. Thus, not(1)$\Rightarrow$not(3). ∎

REMARK 1.3 Using the Taylor approximation at $\mathbb{E}[X]$, we have:

$$\mathbb{E}[u(X)] \simeq u(\mathbb{E}[X]) + (1/2)u"(\mathbb{E}[X])\mathbb{E}[(X - \mathbb{E}[X])^2], \tag{1.4}$$

and also

$$u(\mathbb{E}[X] - \pi[X]) = u(\mathbb{E}[X]) - u'(\mathbb{E}[X])\pi[X]. \tag{1.5}$$

But, by definition, $u(\mathbb{E}[X] - \pi[X]) = \mathbb{E}[u(X)]$. Denote $\sigma_X^2 = \mathbb{E}[(X - \mathbb{E}[X])^2]$. Then:

$$\pi[X] \simeq= -[u"(\mathbb{E}[X])/u'(\mathbb{E}[X])]\sigma_X^2. \tag{1.6}$$

Since, $\sigma_X^2 > 0$, the ratio $-u"(E[X])/u'(\mathbb{E}[X])$ can be also considered as a measure of risk-aversion. Note that it is positive, since the utility function u is increasing and concave. Note also that when $X = V_0 + Y$ with $\mathbb{E}[Y] = 0$, then $\mathbb{E}[X]$ is equal to the initial amount V_0 invested on the market. □

DEFINITION 1.4 *The term $A(x) = -u"(x)/u'(x)$ is called the Arrow-Pratt Measure of Absolute Risk-Aversion (ARA). Another measure allows us to take account of the level of wealth: the ratio $R(x) = -xu"(x)/u'(x)$ which is called the Arrow-Pratt Measure of Relative Risk-Aversion (RRA).*

REMARK 1.4 From the geometrical point of view, note that the curvature of the utility function u is equal to

$$\rho(u)(x) = u"(x)/[1 + u'^2(x)]^{2/3}. \tag{1.7}$$

Taking the absolute value of ρ, we get the risk-aversion $-u"(x)/[1+u'^2(x)]^{2/3}$, which is very close to the ARA measure. □

1.3.2 Standard utility functions

From the previous risk-aversion measures, we can characterize some standard utility functions, and in particular the general class of HARA utilities which are useful to get analytical results.

DEFINITION 1.5 *A utility function u is said to have harmonic absolute risk aversion (HARA) if the inverse of its absolute risk aversion is linear in wealth.*

PROPOSITION 1.1
HARA utility functions u take the following form:

$$u(x) = a\left(b + \frac{x}{c}\right)^{1-c}, \tag{1.8}$$

with u defined on the domain $b + \frac{x}{c} > 0$. *The constant parameters a, b, and c satisfy the condition:* $a(1-c)/c > 0$.
The ARA is given by:

$$A(x) = \left(b + \frac{x}{c}\right)^{-1}, \tag{1.9}$$

which clearly has an inverse linear in wealth x. To ensure that $u' > 0$ *and* $u" < 0$, *it is assumed that* $a\,(1-c)\,/c > 0$.

Usually, three subclasses are distinguished:

- *Constant absolute risk aversion (CARA).* If the parameter c goes to infinity then we obtain $A(x) = A$ constant. In that case, the utility function u takes the form : $u(x) = -\frac{exp[-Ax]}{A}$. Note that the RRA is increasing in wealth.

- *Constant relative risk aversion (CRRA).* If $c = 0$, then $R(x) = c$ constant and, up to a linear transformation, we get:

$$u(x) = \begin{cases} x^{1-c}/(1-c) \; if \; c \neq 1, \\ ln[x] \qquad\qquad if \; c = 1. \end{cases} \tag{1.10}$$

  Note that if $c < 1$, then utility goes from 0 to ∞, and if $c > 1$, then utility goes from $-\infty$ to 0. However, in all cases, this kind of utility function exhibits a *decreasing absolute risk aversion* (DARA).

- *Quadratic utility function.* Consider the case $c = -1$. Then the utility function u is quadratic. Note first that we have to restrict its domain, since u is decreasing on $]b, \infty[$. Second, the ARA is increasing (IARA) with wealth, which unfortunately implies that the risk premium $\pi(.)$ is increasing. Thus, as wealth increases, the unwillingness to take risks increases.

REMARK 1.5 The decreasing absolute risk aversion (DARA) can be characterized by the following property: denote V_0 as the initial amount invested on the market and consider the random payoff X defined by $X = V_0 + Y$ with $\mathbb{E}[Y] = 0$. Then, the risk-premium $\pi_u(V_0, x)$ satisfies: for all $\alpha > 0$, $\pi_u(V_0, x) > \pi_u(V_0 + \alpha, x)$. This property is also equivalent to the existence for all $\alpha > 0$ of a functional denoted by $\Psi_\alpha(.)$, such that: $u(x) = \Psi_\alpha[u(x+\alpha)]$ with $\Psi'_\alpha(.) > 0$ and $\Psi"_\alpha(.) < 0$. ▯

REMARK 1.6 The utility functions CARA and CRRA can be characterized by invariance properties respectively with respect to multiplicative and additive transformations of lotteries (see for example [162]). ▯

1.3.3 Applications to portfolio allocation

Consider the following standard portfolio problem: the investor has an increasing and concave utility function u. He can invest his initial endowment in a risk-free asset M (for example, a monetary asset) and in risky asset S (for example, a stock or a financial index). Denote by r_M and r_S the respective returns of assets M and S. Denote by λ the proportion of initial wealth V_0 invested in the risky asset. Then, the value of the portfolio at maturity is given by:

$$(1-\lambda)V_0(1+r_M)+\lambda V_0(1+r_S)=V_0+\lambda V_0(r_S-r_M). \tag{1.11}$$

Therefore, using the standard assumption $\mathbb{E}[r_S]>r_M$, the expected portfolio return is increasing with respect to the proportion λ. The problem of the investor is to find λ maximizing the expected utility:

$$\max_{\lambda}\mathcal{U}(\lambda)=\mathbb{E}[u(V_0+\lambda V_0(r_S-r_M))]. \tag{1.12}$$

Assuming that u is twice differentiable, the optimum λ^* (if it exists) is deduced from the first-order condition which is given by the implicit relation:

$$\mathcal{U}'(\lambda^*)=\mathbb{E}[(r_S-r_M)u'(V_0+\lambda V_0(r_S-r_M))]=0. \tag{1.13}$$

From the concavity of u, we get:

$$\mathcal{U}"(\lambda^*)=\mathbb{E}[(r_S-r_M)^2u"(V_0+\lambda V_0(r_S-r_M))]\leq 0.$$

Thus, $\mathcal{U}'$ is decreasing. But

$$\mathcal{U}'(0)=u'(V_0)\mathbb{E}[(r_S-r_M))]. \tag{1.14}$$

Thus, from the condition $\mathbb{E}[r_S]>r_M$, we deduce that the portfolio proportion λ^* invested in the risky asset is positive. Note that if $\mathcal{U}'(0)=0$, then $\lambda^*=0$: it is optimal to invest the whole endowment V_0 in the risk-free asset. Thus, the following result is proved:

PROPOSITION 1.2
For the standard portfolio problem with a concave and differentiable utility function, the investor invests a positive proportion in the risky asset if and only if the excess expected return $\mathbb{E}[(r_S-r_M)]$ *is positive.*

REMARK 1.7 The previous result does not depend on the level of the variance of the excess return. Even if the expected excess return is very small, it is optimal to invest part of the portfolio in the risky asset. However, the order of magnitude of λ^* will depend on this variance. Note that if the utility function is not differentiable, then it may happen that $\lambda^*=0$, even if $\mathbb{E}[(r_S-r_M)]>0$. □

Consider the special case of HARA utility function $u(x) = \left(b + \frac{x}{c}\right)^{-c}$. Denote $r_e = (r_S - r_M)$ and λ^{**} the optimal solution when $\left(b + \frac{V_0}{c}\right) = 1$. We have:

$$\begin{aligned} &\mathbb{E}\left[r_e\left(b + \frac{V_0 + \lambda^{**} V_0 \left(b + \frac{V_0}{c}\right) r_e}{c}\right)^{-c}\right] \\ &= \left(b + \frac{V_0}{c}\right)^{-c} \mathbb{E}\left[r_e\left(1 + \frac{\lambda^{**} V_0 r_e}{c}\right)^{-c}\right]. \end{aligned} \tag{1.15}$$

Therefore, since λ^{**} satisfies $\mathbb{E}\left[r_e\left(1 + \frac{\lambda^{**} V_0 r_e}{c}\right)^{-c}\right] = 0$, the optimal solution is $\lambda^* = \lambda^{**}(b + \frac{V_0}{c})$, for any initial endowment V_0. Thus λ^* is a linear function of the wealth V_0. Moreover, when $c \to \infty$, λ^* is independent of wealth.

The hypothesis of Arrow [29] is that individual preferences should be DARA and IRRA (*Increasing relative risk aversion*):

$$\frac{dA(x)}{dx} \leq 0 \text{ and } \frac{dR(x)}{dx} \geq 0. \tag{1.16}$$

The reasoning for DARA is that, for a given risk, wealthy investors are not more risk-averse than poorer ones. IRRA implies that when both wealth and risk increase, then the readiness to bear risk should be reduced. More precisely, for the previous standard portfolio problem with two assets, if the utility function u is twice-differentiable and exhibits DARA and IRRA, then the optimal proportion of initial wealth invested in the risky asset is increasing with wealth; but it increases less than proportionally to the increase in wealth.

REMARK 1.8 When the two assets are risky, another risk aversion measure must be introduced to avoid some paradox, as shown in Ross [434]. For the Ross risk aversion measure, a utility function u is said to display higher risk aversion than the utility function v if there exists a constant $\lambda > 0$ such that for all x and y, we have:

$$\frac{u''}{v''}(x) \geq \lambda \geq \frac{u'}{v'}(y). \tag{1.17}$$

This means that $inf_x u''(x)/v''(x) \geq sup_x u'(x)/v'(x)$. This condition is equivalent to the existence of a constant $a > 0$ and a decreasing concave function G such that for all x in $\mathbb{R}$, $v(x) = au(x) + G(x)$. Finally, it is also equivalent to the following inequality on the risk premia: $\pi_u(V_0, X) \geq \pi_v(V_0, X)$ for any initial wealth V_0 and for any lottery X where $\mathbb{E}[X] = V_0$. $\square$

1.4 Stochastic dominance

How can two random prospects X and Y be ranked? For a given investor, X is prefered to Y if the expected utility of $\mathbb{E}[u(X)]$ is higher than $\mathbb{E}[u(Y)]$. But how can they be compared if the utility is not observable? Besides, we have seen in previous sections how a change in the utility function modifies the risk premium for a given lottery with payoff X. However, how does a change in X modify the risk premium independently of the utility function? The theory of stochastic dominance is introduced to answer these questions. It involves using the probability distributions of random prospects X and Y. In particular, it provides conditions under which $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all utility functions u in a given set. Depending on this set, several stochastic dominance orders are defined.

To simplify, assume that the supports of all random variables are in an interval $[a, b]$. Denote respectively by F_X and F_Y the cumulative distribution functions of X and Y.

DEFINITION 1.6 *X is said to dominate Y according to first-order stochastic dominance ("$X \succeq_1 Y$") if $F_X(w) \leq F_Y(w)$, for all $w \in [a, b]$.*

This definition is consistent with the expected utility, since we have:

PROPOSITION 1.3
$X \succeq_1 Y$ if and only if

$$\mathbb{E}[u(X)] = \int_a^b u(w) dF_X(w) \geq \mathbb{E}[u(Y)] = \int_a^b u(w) dF_Y(w), \tag{1.18}$$

for any utility function u which is monotonically increasing.

PROOF 1) Assume that $X \succeq_1 Y$. Consider any utility function u (monotonically increasing). Integrating by parts, we get:

$$\int_a^b u(w)[dF_X(w) - dF_Y(w)]$$

$$= [u(z)(F_X(w) - F_Y(w))]_a^b - \int_a^b u'(w)[F_X(w) - F_Y(w)]dw.$$

The first term is equal to 0. Moreover, by assumption, $u' > 0$ and $F_X(w) - F_Y(w) \leq 0$. Consequently,

$$\int_a^b u(w)[dF_X(w) - dF_Y(w)] \geq 0,$$

which means that $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$.

2) Conversely, suppose that there exists an w_0 such that $F_X(w_0) > F_Y(w_0)$. By right continuity, there exists a neighborhood $\mathcal{N}(w_0)$ such that $F_X(w) > F_Y(w)$, for all w in $\mathcal{N}(w_0)$. Consider a utility function u which is constant outside $\mathcal{N}(w_0)$ and increasing inside ($u'(w) = 0$ for $w \notin \mathcal{N}(w_0)$ and $u'(w) > 0$ for $w \in \mathcal{N}(w_0)$). Integrating by parts, we get:

$$\int_a^b u(w)[dF_X(w) - dF_Y(w)] = -\int_{\mathcal{N}(w_0)} u'(w)[F_X(w) - F_Y(w)]dw.$$

Since the right side is negative, we deduce $\int_a^b u(w)dF_X(w) < \int_a^b u(w)dF_Y(w)$. Thus "$X \succeq_1 Y$" is false. Consequently, $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for any utility function u (monotonically increasing) implies $X \succeq_1 Y$. $\square$

In the next figure, the random prospects X and Y both stochastically dominate at the first order the random prospect Z. Since for all w, $\mathbb{P}[Z > w]$ is smaller than $\mathbb{P}[X > w]$ and $\mathbb{P}[Y > w]$, the expectation and the expected utility of Z are smaller than those of X and Y.
However, X and Y cannot be compared by this criteria. Indeed, the binary relation $\succeq_1$ is only a partial order on the space of random prospects.

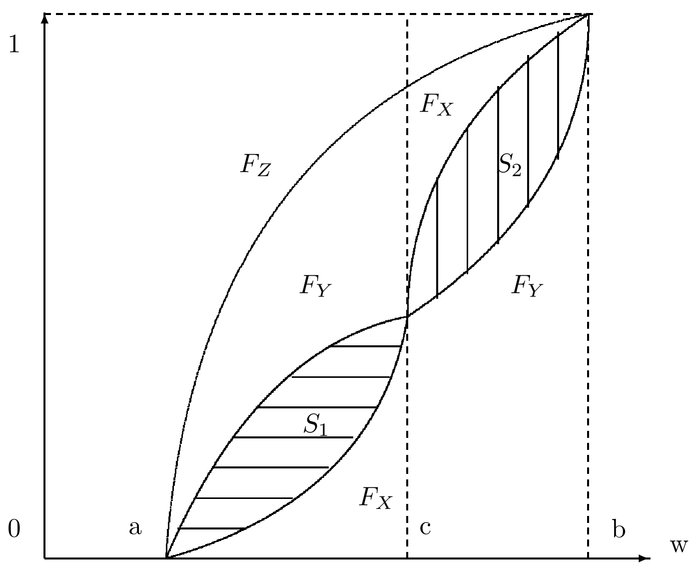


FIGURE 1.2: Stochastic dominance

From Figure (1.2), we see that the random prospect X is "less dispersed"

than Y: the area $S(w) = \int_a^w [F_Y - F_X](s)ds$ between the two curves is always positive (for $w \leq c$, it is in $[0, S_1]$ strictly positive and, assuming $-S_2 \leq S_1$, it is also positive for $w \geq c$).

DEFINITION 1.7 *X is said to dominate Y according to second-order stochastic dominance ("$X \succeq_2 Y$") if $\int_a^w F_X(s)ds \leq \int_a^w F_Y(s)ds$, for all $w \in [a, b]$ or, equivalently, the area $S(w) = \int_a^w [F_Y - F_X](s)ds$ is always positive.*

REMARK 1.9 As mentioned in Hadar and Russel [277], or in Rothschild and Stiglitz [435], this property is equivalent to the fact that all min-utility investors prefer X to Y: for all w_0 in $[a, b]$, we have:

$$\mathbb{E}[Min(X, w_0)] \geq \mathbb{E}[Min(Y, w_0)]. \tag{1.19}$$

Indeed, this latter condition is equivalent to:

$$\int_a^{w_0} w dF_X(w) + w_0(1 - F_X(w_0)) \geq \int_a^{w_0} w dF_Y(w) + w_0(1 - F_Y(w_0)),$$

which is also equivalent to (by integrating by parts)

$$w_0 - \int_a^{w_0} F_X(w)dw \geq w_0 - \int_a^{w_0} F_Y(w)dw,$$

and, finally: $\int_a^{w_0} F_X(w)dz \leq \int_a^{w_0} F_Y(w)dw$, for all w_0 in $[a, b]$. In particular, this property implies that if $X \succeq_2 Y$, then $\mathbb{E}[X] \geq \mathbb{E}[Y]$. Note also that obviously first-order stochastic dominance implies second-order stochastic dominance but not vice-versa. $\square$

A complete caracterization of second-order stochastic dominance is provided by conditions under which $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all utility functions u in a given set:

PROPOSITION 1.4
$X \succeq_2 Y$ if and only if $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$, for any utility function u which is monotonically increasing and concave.

PROOF 1) Assume that $X \succeq_2 Y$. Consider any utility function u increasing concave and twice-differentiable. Integrating by parts, we get:

$$\int_a^b u(w)[dF_X(w) - dF_Y(w)] =$$

$$[u(w)(F_X(w) - F_Y(w))]_a^b - \int_a^b u'(w)[F_X(w) - F_Y(w)]dw.$$

Since the first term is equal to 0, then:

$$\int_a^b u(w)[dF_X(w) - dF_Y(w)] = -\int_a^b u'(w)[F_X(w) - F_Y(w)]dw.$$

Integrating by parts again, we deduce:

$$\int_a^b u(w)[dF_X(w) - dF_Y(w)]$$

$$= \left[-u'(w)\int_a^w (F_X(s) - F_Y(s))ds\right]_a^b + \int_a^b u"(w)\left[\int_a^w (F_X(s) - F_Y(s))ds\right]dw.$$

Recall that by assumption the area $S(w) = \int_a^w [F_Y - F_X](s)ds$ is positive. Then we deduce:

$$\int_a^b u(w)[dF_X(w) - dF_Y(w)]$$

$$= [u'(w)S(w)]_a^b - \int_a^b u"(w)S(w)dw = u'(b)S(b) - \int_a^b u"(w)S(w)dw.$$

Since S, u', and $-u"$ are positive, we get $\int_a^b u(w)[dF_X(w) - dF_Y(w)] \geq 0$, which gives $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$.

2) Conversely, suppose that there exists w_0 such that $S(w_0) < 0$. By continuity, there exists a neighborhood $\mathcal{N}(w_0)$ such that $S(w) < 0$, for all w in $\mathcal{N}(w_0)$. Consider an increasing utility function u which is linear outside $\mathcal{N}(w_0)$ and concave inside ($u"(w) = 0$ for $w \notin \mathcal{N}(w_0)$ and $u"(w) < 0$ for $w \in \mathcal{N}(w_0)$). Then, we get:

$$\int_a^b u(w)[dF_X(w) - dF_Y(w)] = \int_{\mathcal{N}(w_0)} u"(w)S(w)dw.$$

Since the right side is positive, we deduce $\int_a^b u(w)dF_X(w) < \int_a^b u(w)dF_Y(w)$. Thus "$X \succeq_2 Y$" is false. Consequently, $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for any increasing and concave utility function u implies $X \succeq_2 Y$. ∎

REMARK 1.10 (See Hanoch and Lévy [281]). Assume, as in Figure (1.2), that the two curves of F_X and F_Y have only one intersection point. Suppose that there exists c such that:

$$F_X(w) \leq F_Y(w), \text{ for any } w \leq c \text{ and } F_X(w) \geq F_Y(w), \text{ for any } w \geq c.$$

Then:

$$X \succeq_2 Y \text{ if and only if } \mathbb{E}[X] \geq \mathbb{E}[Y]. \tag{1.20}$$

∎

In fact, for two random prospects X and Y with the same returns, the notion of second-order stochastic dominance is equivalent to several definitions of "riskiness," introduced in [277], [281], [435], and [436]. Summing up:

PROPOSITION 1.5
The three following statements are equivalent:

1) Having the same expectation, the expected utility of X is greater than the expected utility of Y for any utility function u which is monotonically increasing and concave:

$$\mathbb{E}[X] = \mathbb{E}[Y] \; and \, \mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]. \tag{1.21}$$

2) Mean-preserving and increasing in spread: the area between the two cumulative distribution functions

$$S(w) = \int_a^w [F_Y - F_X](s)ds \tag{1.22}$$

satisfies: $S(b) = 0$ (i.e. $\mathbb{E}[X] = \mathbb{E}[Y]$), and $S(w)$ is always positive (i.e. $X \succeq_2 Y$).

3) Adding a noise: there exists a random variable ϵ such that $Y = X + \epsilon$ and $\mathbb{E}[\epsilon|X] = 0$.

REMARK 1.11 Statement (3) justifies the diversification. Consider n lotteries with net gains $G_1, ..., G_n$ which are assumed to be independent with the same probability distribution. Consider any feasible strategy $\theta = (\theta_1, ..., \theta_n)$ with weights θ_i such that $\sum_{i=1}^n \theta_i = 1$. Consider the final wealth

$$X = \sum_{i=1}^n \theta_i G_i,$$

associated to the perfect diversification strategy

$$X = \sum_{i=1}^n G_i/n$$

and the final wealth Y associated to any strategy θ. Then the perfect diversification strategy second-order dominates any alternative one: $X \succeq_2 Y$. Indeed, we have:

$$Y = \sum_{i=1}^n \theta_i G_i = X + \sum_{i=1}^n (\theta_i - 1/n) G_i.$$

Since, by symmetry, $E\left[(\theta_i - 1/n)G_i|X\right]$ is independent from i, we deduce:

$$\mathbb{E}\left[\sum_{i=1}^n (\theta_i - 1/n) G_i | X\right] = 0,$$

which implies $X \succeq_2 Y$ by the previous statement (3). $\square$

1.5 Alternative expected utility theory

The key property linked to the expected utility theory is the Independence Axiom, which may fail empirically and can yield some paradoxes, as shown by Allais [19]: consider three outcomes $g_1 = 0$, $g_2 = 100$, and $g_3 = 500$ and two pairs of probability distributions $(\mathbb{P}_1, \mathbb{P}_2)$ and $(\mathbb{Q}_1, \mathbb{Q}_2)$, defined by:

$$\begin{cases} \mathbb{P}_1(g_1) = 0\%, \ \mathbb{P}_1(g_2) = 100\%, \ \mathbb{P}_1(g_3) = 0\%, \\ \mathbb{P}_2(g_1) = 1\%, \ \ \mathbb{P}_2(g_2) = 89\%, \ \mathbb{P}_2(g_3) = 10\%, \\ \\ \mathbb{Q}_1(g_1) = 89\%, \ \mathbb{Q}_1(g_2) = 11\%, \ \ \mathbb{Q}_1(g_3) = 0\%, \\ \mathbb{Q}_1(g_1) = 90\%, \ \ \mathbb{Q}_1(g_2) = 0\%, \ \ \mathbb{Q}_1(g_3) = 10\%. \end{cases}$$

As experimental evidence often shows, usually people prefer lottery $\mathbb{P}_1$ to lottery $\mathbb{P}_2$, and lottery $\mathbb{Q}_2$ to lottery $\mathbb{Q}_1$, while the independence axiom implies that if you prefer $\mathbb{P}_1$ to $\mathbb{P}_2$, then you must prefer $\mathbb{Q}_1$ to $\mathbb{Q}_2$.

Besides some empirical violations of the Independence Axiom, the expected utility theory is a problem for the interpretation of the utility function u, as proved for example in Cohen and Tallon [125]. In fact, the utility function must simultaneously give a representation of the choice among outcomes, and also be the expression of the attitude towards risk. Therefore, for example, an investor who has a decreasing marginal utility u' is necessarily risk-averse.

To argue with this kind of criticism, several responses have been given:

- First, following Marschak [378] or Savage [449], expected utility theory is what "rational" people ought to do under uncertainty, and not necessarily what they actually do. This means that if they were perfectly informed and aware of their decisions, they will behave according to expected utility theory.

- Second, new theories of choice under uncertainty can be introduced to avoid, for example, Allais paradox.

During the 80s, the revision of the expected utility paradigm has been intensely developed by slightly modifying or relaxing the original axioms. Among several proposals are: the *Weighted Utility Theory* of Chew and MacCrimmon [120] that assumes a weaker form of the axiom of independence; the *Prospect Theory* of Kahnemann and Tversky [314] which modifies the axioms much more; the *Non-Linear Expected Utility Theory* of Machina [367]; the *Anticipated Utility Theory* of Quiggin [420]; the *Dual Theory* of Yaari [507]; and the *Regret Theory* introduced by Loomes and Sugden [365].

1.5.1 Weighted utility theory

One of the first theories that was consistent with Allais paradox was the "weighted utility theory", introduced by Chew and MacCrimmon [120] and further developed by Chew [119] and Fishburn [224]. The idea is to apply a transformation on the initial probability.

The basic result of Chew and MacCrimmon yields the following representation of preferences over lotteries $L = \{(\omega_1, p_1), ..., (\omega_m, p_m)\}$:

$$\mathcal{U}(L) = \sum_i u(x_i)\phi(p_i) \text{ with } \phi(p_i) = p_i/[\sum_i v(x_i)p_i], \tag{1.23}$$

where u and v are two different elementary utility functions.

Another approach is the "prospect theory" introduced by Kahnemann and Tversky [314]. Consider the following example where two lotteries are proposed.

Example 1.1

Most people choose A and D. Thus they violate the theory of expected utility

TABLE 1.1: Kahnemann and Tversky example

Problem 1
Assume you are 300 richer than you are today. Choose between: A. The certainty of earning 100 B. 50% probability of winning 200 and 50% of not winning anything
Problem 2
Assume you are 500 richer than today. Choose between: C. A sure loss of 100 D. 50% chance of not losing anything and 50% chance of losing 200

(the independence axiom of the theory). However, in terms of expected utility, the two problems are equivalent: □

TABLE 1.2: Equivalence of the two problems

Problem 1 Case A: 400 with prob=1 Case B: 300 with prob=0.5 or 500 with prob=0.5
Problem 2 Case C: 400 with prob=1 Case D: 300 with prob=0.5 or 500 with prob=0.5

In fact, the available wealth was considered after the choice has been made. Thus, most people behave as risk takers when facing a problem presented in terms of loss (Problem 2), while they behave as risk-averse when the same problem is presented in terms of gain (Problem 1).

This behavioral inconsistency is called the "framing effect," and shows that the mental representation of a choice problem may be crucial.

Kahnemann and Tversky observe that "the preferences observed in the two problems are of particular interest as they violate not only the theory of expected utility, but practically all choice models based on other normative theories."

The idea of the prospect theory is to represent the preferences by means of a function ϕ such that the utility of a lottery

$$L = \{(x_1, p_1), ..., (x_n, p_n)\}$$

is given by:

$$\mathcal{U}(L) = \sum_{i}^{n} u(x_i)\phi(p_i), \tag{1.24}$$

where ϕ is an increasing function defined on $[0, 1]$ with values in $[0, 1]$ and $\phi(0) = 0$, $\phi(1) = 1$.

The function $\phi(\cdot)$ is a transformation of the initial probability and corresponds to a decision weight functional. It allows us to take account of a "certainty effect." For example, if the function ϕ is not left-continuous at 1, then $\phi(p) < p$ maybe in a neighborhood of 1. This is the result of the passage from certitude to uncertainty. Note that the equality $\sum_{i=1}^{n} \phi(p_i) = 1$ may no longer be true.

Using this transformation, the Allais paradox can be solved. Moreover, from experimental observations, Kahneman and Tversky argue that it is necessary to distinguish positive results (gains) from negative ones (losses) from experimental observations. However, the sub-additivity of ϕ which is induced:

$$\forall p_1, p_2 \in]0, 1[\quad \phi(p_1) + \phi(p_2) < \phi(p_1 + p_2), \tag{1.25}$$

may imply the violation of the first-order stochastic dominance, as well as other models with weighted probabilities.

To solve this problem, alternative approaches can be proposed.

1.5.2 Rank dependent expected utility theory

The "Rank Dependent Expected Utility" theory (RDEU) assumes that people consider cumulative distribution functions rather than probabilities themselves. In this framework, it is possible to introduce preference representations that are compatible with the first-order stochastic dominance.

The functional representation of preferences is defined as follows:

DEFINITION 1.8 *For all random variables X and Y which model results or consequences and with values in $[-M, M]$,*

$$X \succ Y \Leftrightarrow V(X) \geq V(Y), \text{ with } V(Z) = \int_{-M}^{M} u(z) d\Phi(F_Z(z)), \tag{1.26}$$

where the function $u(\cdot)$ is continuous and differentiable, non-decreasing and unique up to a non-negative linear, and $\Phi(\cdot)$ is a continuous function, non-decreasing from $[0,1]$ in $[0,1]$. Without loss of generality, it can be assumed that if $\Phi(0) = 0$ and $\Phi(1) = 1$, then $\Phi(\cdot)$ is unique.

Note that for a discrete lottery $L = \{(x_1, p_1), ..., (x_m, p_m)\}$ with

$$x_1 \leq x_2 \leq \cdots \leq x_n,$$

the utility V is given by:

$$\begin{aligned} V(L) &= \textstyle\sum_{i=1}^n u(x_i) \left[\Phi(\sum_{j=1}^i p_j) - \Phi(\sum_{j=1}^{i-1} p_j)\right], \\ &= u(x_1) + \textstyle\sum_{i=2}^n (u(x_i) - u(x_{i-1})) \left[1 - \Phi(\sum_{j=1}^i p_j)\right]. \end{aligned} \tag{1.27}$$

Since the weights $\Phi(\sum_{j=1}^i p_j)$ depend on the ranking of the outcomes x_i, this preference representation is called "rank dependent expected utility." These weights are calculated by first ranking outcomes from the worst to the best then by summing up the utilities weighted by the sequence $(\Phi(\sum_{j=1}^i p_j) - \Phi(\sum_{j=1}^{i-1} p_j))_i$. Thus, it is assumed that an objective probability exists, but individuals transform this given law by using a function of its cdf. Contrary to transformation defined on the pdf itself, this allows us not to violate the first-order stochastic dominance.

REMARK 1.12 The RDEU is a generalization of the expected utility criterion (EU). Indeed, for $\Phi(p) = p, \forall p \in [0,1]$, the functional representation of preferences is given by:

$$V(L) = \sum_{i=1}^n u(x_i) \left[(\sum_{j=1}^i p_j) - (\sum_{j=1}^{i-1} p_j) \right] = \sum_{i=1}^n p_i u(x_i) = EU(\mathcal{L}).$$

If Φ is not the identity but $u(x) = x$, then the RDEU is *the dual theory* of Yaari [507]. ▯

As mentioned in Tallon [488], the RDEU has several advantages:

- Contrary to the EU, the RDEU allows separation of the behavior towards wealth from the behavior towards risk. Therefore, the RDEU is compatible with usual empirical observations which show that individuals under- or overestimate probabilities of random events (*i.e.*, are either pessimistic or optimistic).
- Contrary to the EU, the RDEU allows identification of two notions of risk-aversions: the standard *weak risk-aversion* and the *strong risk-aversion.*

Indeed, in the RDEU framework, these two notions have to be differentiated:

1) The weak risk-aversion of Arrow-Pratt: An individual prefers the expectation of the lottery to the lottery itself:

$$\forall L = (x_i, p_i)_{i=1,\cdots,n}, E(L) = \sum_{i=1}^{n} p_i x_i \succ L. \tag{1.28}$$

In the ES context, it is equivalent to the concavity of the utility function.

2) The strong risk-aversion of Rothschild and Stiglitz ([435] and [436]): This definition is based on the notion of *mean preserving spread.* Consider two random variables X and Y associated respectively to lotteries L_X and L_Y.

DEFINITION 1.9 *Y is said to be a mean preserving spread of X if:*

$$\begin{gathered} \mathbb{E}(L_X) = \mathbb{E}(L_Y) \\ \textit{and} \\ \forall T \in [-M, M], \int_{-M}^{T} Prob\{X < t\}dt \leq \int_{-M}^{T} Prob\{Y < t\}dt. \end{gathered} \tag{1.29}$$

The latter condition corresponds to the second-order stochastic dominance as seen in Proposition 1.5.

DEFINITION 1.10 *A strong risk-averse individual prefers the mean preserving spread Y of X to X itself: $L_Y \succ L_X$.*

REMARK 1.13 These two risk-aversion notions are identical in the ES framework. They correspond to the concavity of the utility function. In the RDEU context, they are distinct.

- Chateauneuf et al. [115] show that an individual satisfying RDEU with a concave utility function $u(\cdot)$ is weakly averse to risk if an only if his transformation function Φ is such that $\Phi(p) \leq p$, $\forall p \in [0,1]$. Furthermore, if $u(x) = 1-(1-x)^n$ with $n \geq 1$, he is weakly averse to risk if and only if his transformation function Φ is such that $\Phi(p) \geq 1-(1-p)^n, \forall p \in [0,1]$.

- Chew, Karni and Safra [121] prove that an individual is strongly averse to risk if and only if his utility function u is concave and his transformation function Φ is convex.

- Tallon [488] proves that strong risk-aversion allows us to give an interpretation of the RDEU: The beliefs of an individual are characterized by a given set of probability distributions and his utility is the infimum of the expectations of his utility on this set. This kind of result is also deduced for characterization of risk measures, as shown in Section 2.1.3 of Chapter 2.

$\square$

Several models have been proposed in the (RDEU) framework, as recalled in what follows.

1.5.2.1 Anticipated utility theory

Quiggin [421] keeps three main properties of the ES theory: the transitivity, the first-order stochastic dominance, and the continuity. He has added the following axiom:

DEFINITION 1.11 *(Weak independence Axiom)*
Consider two lotteries

$$L_X = \{(x_1, p_1), ..., (x_n, p_n)\} \text{ and } L_Y = \{(y_1, p_1), ..., (y_n, p_n)\}$$

such that

$$x_1 \leq ... \leq x_n \text{ and } y_1 \leq ... \leq y_n$$

and

$$\forall i \in \{1, ..., n\}, \ \mathbb{P}[X = x_i] = \mathbb{P}[Y = y_i].$$

Assume that there exists a common value $x_{i_0} = y_{i_0}$. Consider two lotteries $L_{X'}$ and $L_{Y'}$ which are equal respectively to L_X and L_Y, except that x_{i_0} and y_{i_0} are replaced by another common value.
The preference $\succeq$ is weak independent if and only if:

$$L_X \succeq L_Y \Longleftrightarrow L_{X'} \succeq L_{Y'}. \tag{1.30}$$

PROPOSITION 1.6

Consider a lottery $L = \{(x_1, p_1), ..., (x_n, p_n)\}$. Then a functional V which satisfies Quiggin's conditions is given by:

$$V(L) = \sum_{i=1}^{n} u(x_i) \left[\Phi(\sum_{j=1}^{i} p_j) - \Phi(\sum_{j=1}^{i-1} p_j) \right], \tag{1.31}$$

where Φ is non-decreasing on $[0,1]$ to $[0,1]$, and is concave on $[0,\frac{1}{2}]$, ($\Phi(p_i) > p_i$) and convex on $[\frac{1}{2},1]$, ($\Phi(p_i) < p_i$) with $\Phi(\frac{1}{2}) = \frac{1}{2}$ and $\Phi(1) = 1$.

As proposed in Quiggin [420], the function Φ can be chosen as follows:

$$\Phi(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}}, \tag{1.32}$$

with, for example, $\gamma = 0,6$.

REMARK 1.14 Under the previous assumptions, the first-order stochastic dominance is satisfied. Moreover, the Allais paradox is solved. Finally, the model is in accordance with the empirical result of Kahneman and Tversky [314]: individuals weight more events with small probabilities and weight less those with high probabilities. □

REMARK 1.15 For the special case : $u(x) = x$, the RDEU approach is equivalent to the dual theory of Yaari [507]. The implication of such preference representation is that no investor will diversify: in the presence of one riskless asset and one risky asset, either he buys only the riskless one, or only the risky one. However, as shown by Eeckhoudt [185], when both assets are risky, and in the presence of a "background risk" (such as illness, accident, fire...), diversification can be observed. □

Finally, different empirical experiences have shown that individuals do not have the same attitude towards losses and gains: the utility on losses seems to be convex, whereas the utility on gains seems to be concave. The value of each component is computed by taking the expected utility with respect to distortions of the distribution function which may differ for the positive and the negative parts of the distribution. This model can be viewed as a generalization of the standard rank-dependent utility model, where the same distortion function is used for the whole distribution.

This kind of behavior is modelled by the "Cumulative Prospect Theory."

1.5.2.2 Cumulative prospect theory

Tversky and Kahneman [496] have introduced on one hand specific utility functions for losses and gains, on the other hand a transformation function of the cumulative distributions. There exist two functions, w^- and w^+ defined on $[0,1]$, and a utility type function v such that the utility V on the lottery $L = \{(x_1,p_1),...,(x_n,p_n)\}$ with $x_1 < ... < x_m < 0 < x_{m+1} < ... < x_n$ is defined as follows: define Φ^- and Φ^+ by: $\Phi_1^- = w^-(p_1)$ and $\Phi_n^+ = w^+(p_n)$,

$$\begin{aligned}\Phi_i^- &= w^-\left(\textstyle\sum_{j=1}^{i} p_j\right) - w^-\left(\textstyle\sum_{j=1}^{i-1} p_j\right), \forall i \in \{2,...,m\}, \\ \Phi_i^+ &= w^+\left(\textstyle\sum_{j=i}^{n} p_j\right) - w^+\left(\textstyle\sum_{j=i+1}^{n} p_j\right), \forall i \in \{m+1,...,n\}.\end{aligned} \tag{1.33}$$

Then, V is given by: $V(L) = V^-(L) + V^+(L)$ with

$$V^-(L) = \sum_{i=1}^{m} v(x_i)\Phi_i^- \text{ and } V^+(L) = \sum_{i=m+1}^{n} v(x_i)\Phi_i^+. \tag{1.34}$$

When the probability distribution F has a pdf f on $[-M,M]$, and the functions w^- and w^+ have derivatives $w^{-'}$ and $w^{+'}$, then:

$$V(L) = \int_{-M}^{0} v(x)\, w^{-\prime}[F(x)]f(x)dx + \int_{0}^{M} v(x)\, w^{+'}[1-F(x)]f(x)dx.$$

As in Quiggin [420], both functions w^- and w^+ can be chosen as follows: $w(p) = \frac{p^\gamma}{(p^\gamma+(1-p)^\gamma)^{\frac{1}{\gamma}}}$, with, for example, $\gamma^- = 0,69$ and $\gamma^+ = 0,61$.

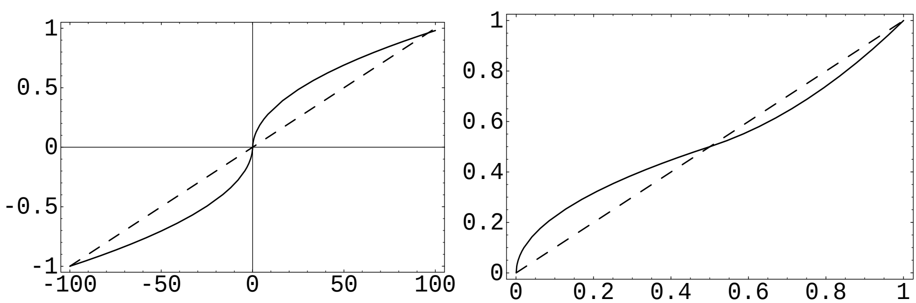


FIGURE 1.3: Kahneman and Tversky functions
The utility function v is convex on losses and concave on gains. The weighting functions w^+ and w^- are above the identical function for small probabilities and under it for large probabilities.

1.5.3 Non-additive expected utility

The "Choquet expected utility" of Schmeidler [454] does not assume that there exists a probability to measure the likelihood of random events. The model is based on the so-called "Choquet Integral" (see Choquet [123]). Consider, for example, a finite probability set Ω equipped with a σ-algebra $\mathcal{F}$.

DEFINITION 1.12 *A Choquet measure (or "capacity") C is a function defined on $\mathcal{F}$ with values in $[0,1]$ satisfying:*

$$\begin{aligned} &C(\emptyset) = 0, \\ &C(\Omega) = 1, \\ &\forall A, B \in \mathcal{F}, A \subset B \Rightarrow C(A) \leq C(B). \end{aligned} \tag{1.35}$$

Note that for mutually exclusive subsets A and B, $C(A \cup B)$ may be smaller or higher than $C(A) + C(B)$.

The Choquet integral, $\int_\Omega u(x) dC(x)$, of a function u (piecewise constant) can be defined as follows:

DEFINITION 1.13 *Let u be defined on the set of outcomes of a lottery $\bar{L} = \{(x_1, A_1), ..., (x_n, A_n)\}$ with $x_1 < ... < x_n$, and where for all i, the event A_i corresponds to the outcome x_i,*

$$\begin{aligned} CEU(\bar{L}) &= u(x_1) + \textstyle\sum_{i=2}^n [(u(x_i) - u(x_{i-1})] C(A_i \cup \cdots \cup A_n). \\ &= \textstyle\sum_{i=1}^n \phi_i u(x_i), \end{aligned} \tag{1.36}$$

where:

$$\forall i = 1, \cdots, n-1, \phi_i = C(A_i \cup \cdots \cup A_n) - C(A_{i+1} \cup \cdots \cup A_n), \phi_n = C(A_n).$$

The Choquet expected utility (CEU)) of an individual is a functional representation of preferences based on some particular axioms.

DEFINITION 1.14 *1) A preference relation $\succeq$ is said to be monotonic if, for any random variables X and Y defined on a probability set Ω:*

$$X(\omega) \geq Y(\omega) \Longrightarrow X \succeq Y. \tag{1.37}$$

2) Two random variables X and Y defined on a probability set Ω are said to be comonotonic if there exist no distinct ω_1 and ω_2 in Ω such that:

$$X(\omega_1) > X(\omega_2) \text{ and } Y(\omega_2) > Y(\omega_1). \tag{1.38}$$

3) A preference relation $\succeq$ is said to be comonotonic independent if, for all pairwise random variables X, Y, and Z, and for any λ in $]0,1[$:

$$X \succ Y \Longrightarrow \lambda X + (1-\lambda) Z \succ \lambda Y + (1-\lambda) Z. \tag{1.39}$$

PROPOSITION 1.7

Assume that the preference relation $\succeq$ satisfies axioms 1 and 2 (see Section (1.1.2)) and monotonicity and comonotonic independence axioms. Then there exists a unique Choquet capacity defined on the σ-algebra $\mathcal{F}$ and an affine real valued function u such that :

$$\bar{L}_X \succeq \bar{L}_Y \Longleftrightarrow \int_\Omega u[X(\omega)]dC(\omega) \geq \int_\Omega u[Y(\omega)]dC(\omega). \tag{1.40}$$

REMARK 1.16 The term $\left[1 - \Phi(\sum_{j=1}^{i} p_j)\right]$ in Definition (1.27) of RDEU corresponds to the term $C(A_i \cup \cdots \cup A_n)$ in Definition (1.36) of CEU. Indeed, for the RDEU, the preference functional is given by

$$V(\mathcal{L}) = \int u_{RDEU}(x)d\Phi(F(x)),$$

while for the CEU, it is defined by

$$V(\bar{L}) = \int u_{CEU}(x)dC(x).$$

□

This preference representation covers problems such as the Ellsberg paradox (see [196]).

1.5.4 Regret theory

Consider the following bet as examined in Lichtenstein and Slovic [356]. Assume that there are two lotteries L_X and L_Y such that:

$$\begin{aligned} &L_X \text{ has outcomes } (A, a) \text{ with probabilities } (p, (1-p)), \\ &L_Y \text{ has outcomes } (B, b) \text{ with probabilities } (q, (1-q)). \end{aligned} \tag{1.41}$$

The money amounts A and B are assumed to be large and a and b are small, possibly negative. The main assumption is that $p > q$ (lottery L_X has a higher probability of a large outcome), and that $B > A$ (lottery L_Y has the highest probability of a large outcome). Thus, individuals who choose lottery L_X face a relatively higher probability of a relatively low gain. Individuals who choose lottery L_Y have a relatively smaller probability of a relatively high gain. For example:

$$\begin{aligned} &L_X \text{ has outcomes } (30, 0) \text{ with probabilities } (p = 90\%, (1-p) = 10\%), \\ &L_Y \text{ has outcomes } (100, 0) \text{ with probabilities } (q = 30\%, (1-q) = 70\%). \end{aligned} \tag{1.42}$$

Note that the expectation of lottery L_X is higher than that of lottery L_Y. Many experiments (see for example [356]) have shown that individuals tend

to choose lottery L_X rather than lottery L_Y. Furthermore, they were willing to sell their right to play lottery L_X for less than their right to play lottery L_Y. Thus, whereas they prefer lottery L_X, they were willing to accept a lower certainty-equivalent amount of money for this lottery (25 for L_X) than they do for the other lottery L_Y (27 for L_Y). This violates the transitivity axiom: indeed, one is indifferent between the certainty-equivalent of a lottery and the lottery itself. Thus, for this example, we have: $U(L_X) = U(25)$ and $U(L_Y) = U(27)$. Since U is increasing, $U(27) > U(25)$, which implies that $U(L_Y) > U(L_X)$. However the empirical choice is such that $U(L_X) > U(L_Y)$, and the preference is not transitive.

Transitivity is often considered as quite rational. However, we can search for preference representation models which can "rationally" explain this "irrationality." Loomes and Sugden [365] proposed a "regret/rejoice" function for pairwise lotteries which contain the outcomes of both the chosen and the foregone lottery. Let L_X and L_Y be two lotteries. If L_X is chosen and L_Y is foregone and the outcome of L_X turns out to be a and the outcome of L_Y turns out to be b, then we can consider, for example, the difference between the (elementary) utilities between the two outcomes, to be a measure of regret, i.e. $r(x, y) = u(x) - u(y)$, which is negative if "regret" and positive if "rejoice." Individuals thus faced with alternative lotteries do not seek to maximize expected utility but rather to minimize expected regret. More generally, the regret theory is defined as follows: Let $R(.)$ be a regret/rejoice function which is assumed to be non-decreasing and such that $R(0) = 0$. Introduce the function Q given by

$$Q[z] = z + R[z] - R[-z]. \tag{1.43}$$

The function Q is non-decreasing and such that $Q[z] = -Q[-z]$. Consider two lotteries $L_X = (x_1, ..., x_n)$ and $L_Y = (y_1, .., y_n)$ with the same probabilities $(p_1, .., p_n)$. The expected rejoice/regret is defined by:

DEFINITION 1.15 *(see [365])*
1) Rejoice/regret of two lotteries:

$$\mathbb{E}(Q(L_X, L_Y)) = \sum_{i=1}^{n} Q[u(x_i) - u(y_j)]p_i. \tag{1.44}$$

2) Preference on lotteries:

$$L_X \succ L_Y \Longleftrightarrow \mathbb{E}(Q(L_X, L_Y)) \geq 0. \tag{1.45}$$

It is obvious that when Q is linear, the regret theory is equivalent to the expected utility criterion. Note that Q is usually assumed to be convex.

1.6 Further reading

Fishburn [223] provides basic results about utility theory for decision-making. In [225], Fishburn recalls the story of expected utility theory based on specific axioms and examines violations of these axioms, which are at the origin of new theories. Gollier [258] provides an overview about expected utility theory and comments on its interest with respect to alternative approaches. Moreover, applications to static portfolio optimization are detailed. Lévy [352] empirically examines the absolute and relative risk aversion to check Arrow's assumption: "investors reveal decreasing absolute risk aversion (DARA) and increasing relative risk aversion (IRRA)." The purpose of the experience is to test these two hypotheses when the individual's wealth varies depending on his/her investment performance. The empirical result is that DARA is indeed strongly supported, but IRRA is rejected.

From the 1980s, and in particular during the 1990s, new theories based on alternative axioms have been introduced such that the Allais hypothesis would emerge as a result. The debate about expected or non-expected utility has not yet ended. A comprehensive and detailed survey of the literature on alternative expected utility theories is found in Fishburn [226]. Among them, the non-linear expected utility of Machina [367] appealed to the notion of "local expected utility." A generalization of rank dependent expected utility is introduced in Segal ([457],[458]): there is no strict separation between behavior towards outcomes and their probabilities (joint function of outcomes and cdf).

Maccheroni et al. [369] introduce dynamic variational preferences. They generalize the multiple priors preferences of Gilboa and Schmeidler [254] who model ambiguity averse agents. They provide conditions under which dynamic variational preferences are time consistent and have a recursive representation.

Experiments have been conducted to check that individual choices do not satisfy the linearity property with respect to probabilities. Schoemaker [455] provides several examples showing that the expected utility property is often violated: the choice process is examined from economic, decision theoretic, and psychological perspectives. Multiple factors are examined which may induce variance in risk-bearing: including portfolio constraints and market incompleteness. In the RDEU framework, the major difficulty is to estimate separately the utility function and the transformation function. Wakker and Deneffe [501] provide a method to dissociate these two estimations. In [160] and [161], another approach is proposed: the idea is to test the invariance of choices with respect to a given transformation (for example, additive or multiplicative). This allows a characterization of the utility functions, independently of the weighting of the utility. Jaffray [295] and Cohen and Tallon

[125] examine the individual's behavior when he focuses simultaneously on the worst and best outcomes. They propose special weighting of the utilities of these outcomes.

Many recent studies are devoted to behavorial finance and its application to portfolio management. For example, in Benzion and Yagil [53], portfolio choices are investigated experimentally. Other consequences for financial markets are analyzed, for instance in De Bondt and Wolff [151]. In Lévy and Lévy [353], it is shown that, when diversification between assets are allowed, mean-variance and prospect theory determine almost the same efficient sets, while prospect theory supposes that investors' choices are based on change of wealth rather than on total wealth, and that objectice probabilities are subjectively distorted.

Another approach, which has not yet been fully applied in financial modelling, is based on *the discrete choice theory* as detailed by MacFadden [366]. The idea is that individual preferences are random utility functions: one component is observable and deterministic. The other component, which is not observable, takes account of imperfections such as lack of information, individuals' heterogeneity, *etc.* The Logit multinomial model is frequently used to model the uncertainty about the utility function.

Chapter 2

Risk measures

Recent years have seen increasing development of new tools for risk management analysis. Many souces of risk have been identified, such as market risk, credit risk, counterparty default, liquidity risk, operational risk and others. One of the main problems concerning the evaluation and optimization of risk exposure is the choice of "good" risk measures. Value-at-Risk has been introduced for bank regulation purposes. Nevertheless, due to some of its deficiencies, other risk measures have been proposed. But what axioms must be imposed to determine a "rational" risk measure?

2.1 Coherent and convex risk measures

In portfolio theory, many risk measures alternative to the variance have been introduced. As noted by Giacometti and Ortobelli [252], we can distinguish two kinds of risk measures:

- First, the *dispersion* measures: these functions of risks X are increasing, positive, and positively homogeneous. Among them: standard deviation and mean-absolute deviation. Note that, if $X > 0$ and $\lambda > 1$, then

  $$\rho(\lambda X) = \lambda \rho(X) > \rho(X).$$

  Thus, they are not consistent with the first-order stochastic dominance.

- Second, the *safety* measures: these were introduced in portfolio theory by Roy [437], Telser [490], and Kataoka [324]. The safety-first rules are based on risk measures which involve the probability that the portfolio return falls under a given level. These kinds of measures are consistent with the first-order stochastic dominance.

Among the safety-first measures, a special class has been emphasized: the *coherent* risk measures.

2.1.1 Coherent risk measures

In their seminal contribution, Artzner, Delbaen, Eber, and Heath [31] introduce axioms that are based on regulation of private banks by a central bank. Such risk measures determine the minimal amount of capital that must be added to make the future value of a position acceptable. Consider the set Ω of states of nature. This set may describe, for example, the possible values of all asset prices. The probabilities of each state ω may be unknown. Let $X(\omega)$ be the discounted future net worth of the position for each "scenario" ω. Let $\mathcal{X}$ be the set of all risks; that is a given set of real random variables defined on Ω. The set $\mathcal{X}$ is a linear space which may be, for instance, the set of all bounded random variables.

Assume that there exists a reference asset of a constant total return r. Recall the axiomatics of Artzner et al. for a risk measure ρ defined on the set $\mathcal{X}$:

- **Axiom T.** *Translation invariance:* for all X in $\mathcal{X}$ and all real numbers α, we have
$$\rho(X + \alpha.r) = \rho(X) - \alpha.$$
This means that if the sure amount α is initially invested in the reference asset, then the variation of the risk measure is equal to α itself. This property is quite in accordance with a monetary interpretation of the measure ρ. Note that in particular, we have:
$$\rho(X + \rho(X).r) = 0.$$

- **Axiom S.** *Subadditivity:* for all X_1 and X_2 in $\mathcal{X}$:
$$\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2).$$
In particular, this property means that "a merger does not create extra risk." Note also that, for a global risk $X_1 + X_2$, the amount $\rho(X_1) + \rho(X_2)$ is sufficient to guarantee the position. This property does not allow reduction of risk by dividing the total position into smaller ones, which is highly desirable for regulation purposes, but does not satisfy the diversification principle.

- **Axiom PH.** *Positive homogeneity:* for all X in $\mathcal{X}$ and for all $\lambda \geq 0$,
$$\rho(\lambda X) = \lambda \rho(X).$$
This property implies that the risk measure is a linear function of the size of the position. It requires that there is no liquidity risk.

- **Axiom M.** *Monotonicity:* for all X_1 and X_2 in $\mathcal{X}$:
$$X_1 \leq X_2 \Longrightarrow \rho(X_1) \geq \rho(X_2).$$

- **Axiom R.** *Relevance:* for all X in $\mathcal{X}$ with $X \leq 0$ and $X \neq 0$,

$$\rho(X) > 0.$$

  This means that if there actually exists a risk, it must be taken into account.

DEFINITION 2.1 *A risk measure satisfying the axioms of translation invariance, subadditivity, positive homogeneity, and monotonicity is called coherent.*

2.1.2 Convex risk measures

When the risk measure is not assumed to have variations proportional to the risk variations themselves, the positive homogeneity is no longer satisfied. Alternative axioms can be proposed. This leads to the notion of convexity, as introduced in Heath [286] for finite sample spaces ("weakly coherent risk measures") and Föllmer and Schied [234] for general spaces (see also [236]).

- **Axiom C** *Convexity:* for all X_1 and X_2 in $\mathcal{X}$ and for all $0 \leq \lambda \leq 1$,

$$\rho(\lambda X_1 + (1-\lambda)X_2) \leq \lambda\rho(X_1) + (1-\lambda)\rho(X_2).$$

DEFINITION 2.2 *A risk measure satisfying the axioms of translation invariance, monotonicity, and convexity is called convex.*

PROPOSITION 2.1
A convex risk measure is coherent if it satisfies the positive homogeneity. Note also that the positive homogeneity and the subadditivity implies the convexity.

As mentioned in Artzner et al. [31], a class A_ρ of "acceptable positions" can be associated to each risk measure ρ. This class, called the *acceptance set* of ρ, contains all positions X which do not induce positive risk:

$$A_\rho = \{X \in \mathcal{X} | \rho(X) \leq 0\}. \tag{2.1}$$

Conversely, given a class A in $\mathcal{X}$ of acceptable positions, a risk measure ρ_A can be defined by:

$$\rho_A(X) = inf\{m \in \mathbb{R}\,|m + X{\in}A\,\}. \tag{2.2}$$

The following proposition (see [236] and [238]) indicates the relations between convex risk measures and the corresponding acceptance sets.

PROPOSITION 2.2
If ρ is a convex risk measure with acceptance set A_ρ, then $\rho_{A_\rho} = \rho$. Besides, the acceptance set $A = A_\rho$ has the following properties:
1) The set A is non-empty and convex.
2) If $X \in A$ and $Y \in \mathcal{X}$, then $Y \geq X$ implies $Y \in A$.
3) If the risk measure ρ is coherent, then A is a convex cone.
Conversely, if A is a non-empty convex subset of $\mathcal{X}$ which satisfies property (2) and such that the corresponding functional ρ_A satisfies $\rho_A(0) > -\infty$, then:
1) ρ_A is a convex measure of risk.
2) If A is a cone then ρ_A is a coherent measure of risk.

2.1.3 Representation of risk measures

Assume that $\mathcal{X}$ is the linear space of all bounded measurable functions on a measurable space $(\Omega, \mathcal{F})$. Denote by $\mathcal{M}_1 = \mathcal{M}_1(\Omega, \mathcal{F})$ the class of all probabilities on $(\Omega, \mathcal{F})$. Denote also by $\mathcal{M}_{1,f} = \mathcal{M}_{1,f}(\Omega, \mathcal{F})$ the class of all finitely additive and non-negative functions $\mathbb{Q}$ on $\mathcal{F}$ such that $\mathbb{Q}(\Omega) = 1$. Then a characterization of coherent risk measures can be deduced (see [30] and[31] for finite probability spaces, and [154] for general spaces):

PROPOSITION 2.3
A functional ρ is a coherent measure of risk if and only if there exists a subset $\mathcal{Q}$ of $\mathcal{M}_{1,f}$ such that:

$$\rho(X) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[-X], \; X \in \mathcal{X}. \tag{2.3}$$

Besides, $\mathcal{Q}$ can be chosen such that it is convex and the supremum is attained.

A similar representation result is obtained in [237] for convex measures of risk. Let $\alpha : \mathcal{M}_1(\Omega, \mathcal{F}) \to \mathbb{R} \cup \{\infty\}$ be a functional which is bounded from below and not identically equal to ∞. Define the measure of risk ρ by:

$$\rho(X) = \sup_{\mathbb{Q} \in \mathcal{M}_{1,f}} \left(\mathbb{E}_{\mathbb{Q}}[-X] - \alpha(\mathbb{Q})\right). \tag{2.4}$$

The measure ρ associated to the function α is convex.

DEFINITION 2.3 *The functional α is called a penalty function for the risk measure ρ defined on $\mathcal{M}_{1,f}$.*

PROPOSITION 2.4
Any convex measure of risk ρ on $\mathcal{X}$ has the following form:

$$\rho(X) = \max_{\mathbb{Q} \in \mathcal{M}_{1,f}} \left(\mathbb{E}_{\mathbb{Q}}[-X] - \alpha_{min}(\mathbb{Q})\right), \; X \in \mathcal{X}, \tag{2.5}$$

where the penalty function α_{min} is given by:

$$\alpha_{min}(\mathbb{Q}) = \sup_{Y \in A_\rho} \mathbb{E}_{\mathbb{Q}}[-Y], \; for \; \mathbb{Q} \in \mathcal{M}_{1,f}. \tag{2.6}$$

Additionally, the minimal penalty function α_{min} represents the risk measure ρ: for any penalty function α satisfying relation (2.5), $\alpha(\mathbb{Q}) \geq \alpha_{min}(\mathbb{Q})$, for all $\mathbb{Q} \in \mathcal{M}_{1,f}$.

REMARK 2.1 - The minimal penalty function α_{min} of a coherent measure of risk ρ takes only the values 0 and ∞ and:

$$\rho(X) = \max_{\mathbb{Q} \in \mathcal{Q}_{max}} \mathbb{E}_{\mathbb{Q}}[-X], X \in \mathcal{X},$$

for the convex set

$$\mathcal{Q}_{max} = \{\mathbb{Q} \in \mathcal{M}_{1,f} | \alpha_{min}(\mathbb{Q}) = 0\},$$

which is the largest set for which the representation (2.3) holds.
- If $\mathcal{Q}$ is the set of all probability measures on $(\Omega, \mathcal{F})$, then the coherent risk measure induced by $\mathcal{Q}$ is given by *the worst case measure*:

$$\rho_{max}(X) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[-X] = -\inf_{\omega \in \Omega} X(\omega), \text{ for all } X \in \mathcal{X}.$$

- Using continuity argument such as, if $X_n \searrow X$ then $\rho(X_n) \nearrow \rho(X)$, the measure of risk ρ can be represented as the maximum:

$$\rho(X) = \max_{\mathbb{Q} \in \mathcal{M}_1} \left(\mathbb{E}_{\mathbb{Q}}[-X] - \alpha_{max}(\mathbb{Q})\right), X \in \mathcal{X},$$

which may be too restrictive (see [237]). ▯

2.1.4 Risk measures and utility

As detailed in Chapter 1, under some specific assumptions, the investor's preference on the possible positions X can be represented by a utility function U. Nevertheless, the expected utility criterion, which assumes in particular that one single probability measure is determined, may be not appropriate to model the investor's decision under uncertainty. Maybe the investor has in mind a larger class of measures of occurrence of scenarios, represented by elements of $\mathcal{M}_{1,f}$. Then, he may consider the worst case for the expected losses of his investments. In this framework, under a set of axioms for his preference order (see [253] and [254]), there exists a Savage representation of U associated to a utility function u defined on the outcomes such that:

$$U(X) = \inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[X], \tag{2.7}$$

where $\mathcal{Q}$ is a set of probability measures on $(\Omega, \mathcal{F})$. A coherent measure of risk ρ can be associated to this functional U from the relation:

$$U(X) = -\rho(u(X)), \text{ for all } X \in \mathcal{X}. \tag{2.8}$$

As detailed in [237] (see also [303]), another approach is to associate a *loss functional* L to the utility U by letting $L = -U$. Then, there exists a *loss function* l, convex and increasing, defined by $l(x) = -u(-x)$ and such that:

$$L(X) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[l(-X)]. \tag{2.9}$$

A position X is acceptable if the loss functional $L(X)$ is not higher than a given reference level x_r. Then, a convex class of acceptable positions is deduced:

$$A_L = \{X \in \mathcal{X} | L(X) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[l(-X)] \leq x_r\}.$$

The set A defines a convex measure of risk ρ_L with the representation:

$$\rho_L(X) = \sup_{\mathbb{Q}_i \in \mathcal{M}_1} \left(\mathbb{E}_{\mathbb{Q}}[-X] - \alpha_L(\mathbb{Q})\right).$$

Then, the penalty function α_L has to be determined. As proved in [238], a general expression can be provided by using the Fenchel-Legendre transform l^* of the loss function l which is defined by:

$$l^*(y) = \sup_{x \in \mathbb{R}} \left[yx - l(x)\right].$$

PROPOSITION 2.5
The convex risk measure ρ_L associated to the acceptance set A_L has a penalty function α_L given by:

$$\alpha_L(P) = \inf_{\lambda > 0} \frac{1}{\lambda} \left(x_r + \inf_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}} \left[l^* \left(\lambda \frac{d\mathbb{P}}{d\mathbb{Q}} \right) \right] \right), \tag{2.10}$$

where $\frac{d\mathbb{P}}{d\mathbb{Q}}$ is a generalized density in the sense of Lebesgue decomposition: $\alpha_L(P) < \infty$ only if P is absolutely continuous with respect to at least some $\mathbb{Q} \in \mathcal{Q}$ (notation: $\mathbb{P} \ll \mathbb{Q}$).

Example 2.1
For the exponential loss function $l(x) = e^x$ and $x_r = 1$, the penalty function α_L is given by:

$$\alpha_L(\mathbb{P}) = \inf_{\mathbb{Q} \in \mathcal{Q}} H(\mathbb{P}/\mathbb{Q}), \tag{2.11}$$

where $H(\mathbb{P}/\mathbb{Q})$ denotes the relative entropy of $\mathbb{P}$ with respect to $\mathbb{Q}$:

$$H(\mathbb{P}/\mathbb{Q}) = \begin{cases} \mathbb{E}_{\mathbb{Q}} \left[\frac{d\mathbb{P}}{d\mathbb{Q}} log \frac{d\mathbb{P}}{d\mathbb{Q}} \right] & \text{if } \mathbb{P} \ll \mathbb{Q}, \\ +\infty & \text{otherwise.} \end{cases}$$

$\square$

2.1.5 Dynamic risk measures

In Cvitanic and Karatzas [140], the idea of risk measures is introduced in a dynamic setting. They assume that a complete financial market is given and they define the risk of a position as the highest expected shortfall under some set of probability measures. The corresponding risk measure is given by:

$$\rho(X)_{V_0} = \sup_{\mathbb{Q}\in\mathcal{Q}} \inf_{\theta\in\Theta(V_0)} \mathbb{E}_{\mathbb{Q}} \left[(X - V_T^{\theta})\right]^+, \tag{2.12}$$

where $\mathcal{Q}$ is a given set of probability measures ("the possible scenarios"), $\Theta(V_0)$ is the class of all admissible portfolios with initial endowment V_0, and V_T^{θ} is the value of the portfolio θ at maturity T.

Thus, as mentioned in Frittelli [243], the dynamic property of such a measure is due to the possible portfolio rebalancing, and not to a dynamic value of the measure itself. In Wang [502], the risk measure is introduced in a dynamic way. The class of "likelihood-based risk measures" is introduced to take account of dynamic readjustments (for discrete time and finite sample spaces). Further results can be found in [32], where two different dynamic measures of risk are proposed.

In Riedel [425], the class of coherent risk measures is extended to the dynamic framework by considering a *predictable* translation invariance to take new information into account, and also by introducing a *dynamic consistency* to avoid possible contradictions between judgements over time. This leads to the following representation:

Consider a sequence of time periods $t = 0, ..., T$ and a finite set Ω of states of the world. Suppose that there exists a sequence of random variables $(Z_t)_t$ which reveals the information at time t.

Denote by $(\mathcal{F}_t)$ the filtration generated by the process Z:

$$\mathcal{F}_t = \sigma(Z_1, ..., Z_t). \tag{2.13}$$

A position $D = (D_t)_t$ is a $\mathcal{F}_t$-adapted process, which corresponds to a sequence of random payments at any time t. It is assumed that there exists an exogenous interest rate r. Then, under the axioms given in [425], the dynamic risk measure assigns to the sequence D of payments the risk:

$$\rho_t(D) = \max_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}} \left[-\sum_{s=t}^{T} \frac{D_s}{(1+r)^{T-s}} |\mathcal{F}_t \right]. \tag{2.14}$$

The notion of dynamic risk measures can be defined as follows (see [243]):

- First, a whole process $(\rho_t)_t$ must be introduced to represent the risk of the position at any time, conditionally to the information available at any time t along the period $[0, T]$.
- Second, some boundary conditions must be satisfied. At time 0, the risk measure ρ_0 is a static risk measure as defined in previous sections. Additionally, at maturity T, ρ_T is the opposite of the worth of the financial position.

Let $(\mathcal{F}_t)_t$ be a filtration defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Denote by $\mathbb{L}^p = \mathbb{L}^p(\Omega, \mathcal{F}_T, \mathbb{P})$ the space of all real-valued, $\mathcal{F}_T$-measurable, and p-integrable random variables, and by $\mathbb{L}_t^0 = \mathbb{L}^0(\Omega, \mathcal{F}_t, \mathbb{P})$ the space of all random variables defined on $(\Omega, \mathcal{F}_t, \mathbb{P})$.

DEFINITION 2.4 *A dynamic risk measure is a process $(\rho_t)_t$ such that:*
i) $\rho_t : L^p \to \mathbb{L}_t^0$, for all $t \in [0, T]$;
ii) ρ_0 is a static measure; and,
iii) $\rho_T(X) = -X$, $\mathbb{P} - a.s.$, for all $X \in L^p$.

Most of the "rational" dynamic axioms that can be imposed on the process $(\rho_t)_t$ are similar to those in the static case.

Axiom T. *Translation invariance:* $\forall t \in [0, T]$, $\forall a \in \mathbb{L}^p$ and $\mathcal{F}_t$-measurable, $\forall X \in \mathbb{L}^p$,

$$\rho_t(X + a) = \rho_t(X) - a, \ \mathbb{P} - a.s.$$

Note that at time t, the random variable a can be considered as a constant since it is observable given the information $\mathcal{F}_t$.

Axiom S. *Subadditivity:* $\forall t \in [0, T]$, $\forall X_1, X_2 \in \mathbb{L}^p$,

$$\rho_t(X_1 + X_2) \leq \rho_t(X_1) + \rho_t(X_2), \ \mathbb{P} - a.s.$$

Axiom PH. *Positive homogeneity:* $\forall t \in [0, T]$, $\forall X \in \mathbb{L}^p$ and $\forall \lambda \geq 0$,

$$\rho_t(\lambda X) = \lambda \rho_t(X), \ \mathbb{P} - a.s.$$

Axiom M. *Monotonicity:* $\forall X_1, X_2 \in \mathbb{L}^p$,

$$X_1 \leq X_2 \Longrightarrow \forall t \in [0, T], \ \rho_t(X_1) \geq \rho_t(X_2), \ \mathbb{P} - a.s.$$

Axiom Ct. *Constancy:* $\forall c \in \mathbb{R}$, $\forall t \in [0, T]$,

$$\rho_t(c) = -c, \ \mathbb{P} - a.s.$$

Axiom P. *Positivity:* $\forall X \in \mathbb{L}^p$,

$$X \geq 0 \Longrightarrow \forall t \in [0, T], \rho_t(X) \leq \rho_t(0), \ \mathbb{P} - a.s.$$

Axiom C. *Convexity:* $\forall t \in [0,T]$, $\forall X_1, X_2 \in \mathbb{L}^p$, and $\forall \lambda$ such that $0 \leq \lambda \leq 1$,

$$\rho_t(\lambda X_1 + (1-\lambda)X_2) \leq \lambda \rho_t(X_1) + (1-\lambda)\rho_t(X_2).$$

The following definitions are proposed for coherent and convex risk measures.

DEFINITION 2.5 *(see [243])*
i) A dynamic risk measure $(\rho_t)_t$ is called convex if it satisfies axiom (C) and $\rho_t(0) = 0$.
ii) A dynamic risk measure $(\rho_t)_t$ is called coherent if it satisfies axioms (T), (S), (PH), and (P).
iii) A dynamic risk measure $(\rho_t)_t$ is said to be time-consistent if it satisfies the following axiom: $\forall t \in [0,T]$, $\forall X \in \mathbb{L}^p$, $\forall A \in \mathcal{F}_t$,

$$\rho_0(XI_A) = \rho_0(-\rho_t(X)I_A).$$

This time-consistency condition is the condition of the "filtration-consistency" of Coquet et al. [129] adapted to the risk measure framework. It is linked to the recursivity property defined in Artzner *et al.* [32] (not their time-consistency definition).

There exists mainly two ways to provide dynamic risk measures:
- First, by using robust representations as in the static case, as shown in ([32], [33]).
- Second, by relying on dynamic risk measures to backward stochastic differential equations (BSDE) through the notion of the "conditional g-expectation," as introduced in Peng [403].

The first approach can be summarized into the two following results:

Dynamic convex risk measure: Let $\mathcal{Q}$ be a convex set of $\mathbb{P}$-absolutely continuous probability measures defined on $(\Omega, \mathcal{F}_t)$. For any $t \in [0,T]$, let $\alpha_t : \mathcal{Q} \to \mathbb{R}$ be a convex functional such that $\inf_{\mathbb{Q}\in\mathcal{Q}} \alpha_t(\mathbb{Q}) = 0$. Then, the process $(\rho_t)_t$ is defined by: $\forall t \in [0,T]$, $\forall X \in \mathbb{L}^p$,

$$\rho_t(X) = \text{ess. sup}_{\mathbb{Q}\in\mathcal{Q}} \left(\mathbb{E}_{\mathbb{Q}}[-X|\mathcal{F}_t] - \alpha_t(\mathbb{Q})\right). \tag{2.15}$$

This relation is the dynamic version of relation (2.5): any such process $(\rho_t)_t$ is a dynamic convex risk measure. In addition, it satisfies the axioms (T), (Ct), and (P) (*i.e.*, translation invariance, constancy and positivity).

Dynamic coherent risk measure: Let $\mathcal{Q}$ be a convex set of $\mathbb{P}$-absolutely continuous probability measures defined on $(\Omega, \mathcal{F}_t)$. Then, the process $(\rho_t)_t$ is defined by: $\forall t \in [0,T]$, $\forall X \in \mathbb{L}^p$,

$$\rho_t(X) = \text{ess. sup}_{\mathbb{Q}\in\mathcal{Q}} \left(\mathbb{E}_{\mathbb{Q}}[-X|\mathcal{F}_t]\right). \tag{2.16}$$

This is the dynamic version of relation (2.3): any such process $(\rho_t)_t$ is a dynamic coherent risk measure.

The second approach is based on the following BSDE:

Consider a standard d-dimensional Brownian motion $(B_t)_t$, and denote $(\mathcal{F}_t)_t$ the augmented filtration generated by B.
Denote by $\mathbb{L}^2_{\mathcal{F}} = \mathbb{L}^2_{\mathcal{F}}(T, \mathbb{R}^n)$ the space of all $\mathbb{R}^n$-valued, adapted processes θ such that:

$$E\left[\int_0^T \|\theta_t\|_n^2 dt\right] > \infty, \tag{2.17}$$

where $\|.\|_n$ stands for the Euclidean norm on $\mathbb{R}^n$.

Let $g : [0,T] \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ be a function which satisfies usual conditions to guarantee existence and uniqueness of the solution (Y, Z) of the following BSDE (see [403] or [129]):

$$-dY_t = g(t, Y_t, Z_t)dt - Z_t dB_t,\ 0 \le t \le T \text{ and } Y_T = X.$$

The notion of g-expectation introduced in Peng [403] is as follows:

DEFINITION 2.6 *For any $X \in \mathbb{L}^2$ and for all $t \in [0,T]$, the conditional g-expectation of X under $\mathcal{F}_t$, denoted by $\mathcal{E}_g[X|\mathcal{F}_t]$ is defined by:*

$$\mathcal{E}_g[X|\mathcal{F}_t] = Y_t, \tag{2.18}$$

where Y is the first component of the solution of the previous BSDE with value X at maturity T. For $t = 0$, $\mathcal{E}_g[X] = Y_0$ is called the g-expectation.

Then, a risk measure can be defined as a family of maps from to $\mathbb{L}^2$ to $\mathbb{L}^2_{\mathcal{F}}$. Consider the process $(\rho_t)_t$ given by: for all $X \in \mathbb{L}^2$,

$$\rho_t(X) = \mathcal{E}_g[-X|\mathcal{F}_t]. \tag{2.19}$$

From the representation of risk measures by conditional g-expectations, sufficient conditions of convexity and coherence can be deduced (see [243]):

PROPOSITION 2.6
i) If the functional g is convex in $(y, z) \in (\mathbb{R} \times \mathbb{R}^d)$, then the process $(\rho_t)_t$ defined in (2.19) is a convex dynamic risk measure. It is also time-consistent and satisfies axioms (T), (P), and (Ct).
ii) If the functional g is sublinear in $(y, z) \in (\mathbb{R}, \mathbb{R}^d)$, then the process $(\rho_t)_t$ defined in (2.19) is a coherent dynamic risk measure. It is also time-consistent.

Conversely, sufficient conditions can be given to prove that a given risk measure is associated to a g-expectation.

For this purpose, recall the definition of $\mathcal{E}_\mu$-dominance given in [129]: the measure ρ_0 is $\mathcal{E}_\mu$-dominated if for any X_1 and X_2 in $\mathbb{L}^2$,

$$\rho_0(X_1 + X_2) - \rho_0(X_1) \leq \mathcal{E}_{g_\mu}[-Y], \text{ for some } g_\mu = \mu|z|, \text{ with } \mu > 0. \quad (2.20)$$

Assume that $d = 1$. Then the following result can be deduced:

PROPOSITION 2.7
Consider a dynamic convex risk measure $(\rho_t)_t$ on $\mathbb{L}^2$ satisfying translation invariance (T) and monotonicity (M).
1) If:
i) $(\rho_t)_t$ is time-consistent;
ii) ρ_0 is strictly monotone, i.e.:

$$X_1 \geq X_2 \textit{ and } \rho_0(X_1) = \rho_0(X_2) \Longleftrightarrow X_1 = X_2; \quad (2.21)$$

iii) and, ρ_0 is $\mathcal{E}_\mu$-dominated,

then there exists a unique functional $g : \Omega \times [0,T] \times \mathbb{R} \to \mathbb{R}$, independent of y, satisfying usual conditions, such that $|g(t,z)| \leq \mu|z|$ and associated to ρ by: for all X in $\mathbb{L}^2$,

$$\rho_0(X) = \mathcal{E}_g(-X) \textit{ and } \rho_t(X) = \mathcal{E}_g[-X|\mathcal{F}_t]. \quad (2.22)$$

Besides, if g is $\mathbb{P}$-a.s. continuous in t for all $z \in \mathbb{R}$ then g is convex in z.

2) Consider a dynamic coherent risk measure $(\rho_t)_t$ on $\mathbb{L}^2$ satisfying properties (i), (ii), and (iii). Then there exists a unique functional g, independent of y, satisfying (2.22). Also, if g is $\mathbb{P}$-a.s. continuous in t for all $z \in \mathbb{R}$, then g is sublinear in z.

In [44], a general axiomatic of dynamic convex risk measures is related to the notions of consistent convex price systems and non-linear expectations introduced in El Karoui and Quenez [194] and Peng [403]. Then their dynamic risk measures (so-called "market risk measures") are also related to quadratic BSDE.

2.2 Standard risk measures

2.2.1 Value-at-Risk

The need to introduce new specific risk measures, alternative to dispersion measures such as standard deviation, has been implied by the regulation of financial institutions and the development of risk management. The Value-at-Risk (VaR) has been the first attempt to provide a quantitative tool to determine capital requirements (see for example JP Morgan [393]), and to better take account of fat-tailed and non-normal returns; for example, when volatilities are random and possible jumps may occur, when financial options are involved in the position, when default-risks are not negligible, when cross-dependence between assets are complex, *etc.*

2.2.1.1 Definition of the VaR

To illustrate the VaR concept, consider the following model:

Example 2.2
Let $a = (a_1, ..., a_N)'$ be a portfolio allocation, where a_i denotes the allocation invested on the i-th financial asset. Denote by $S_{i,t}$ the price of the asset i at time t. Then, the portfolio value $V(a)$ is given by:

$$V_t(a_t) = \sum_{i=1}^{N} a_{i,t} S_{i,t} = a'.S_t.$$

Then, its change between dates t and $t+1$ is equal to:

$$\Delta V_{t+1}(a_t) = (V_{t+1} - V_t)(a_t) = \sum_{i=1}^{N} a_{i,t}(S_{i,t+1} - S_{i,t}) = a'.\Delta S_{t+1}.$$

Assume that the future asset prices $S_t = (S_{1,t}, ..., S_{N,t})$ has a continuous conditional distribution $\mathbb{P}_{S_t,t}$ given the information at time t. Then, for a loss probability level α, with usual values between 1% and 5%, the Value-at-Risk $VaR_{t,\alpha}(a_t)$ is defined by:

$$\mathbb{P}_{S_t,t}\left[(V_{t+1} - V_t)(a_t) + VaR_{t,\alpha}(a_t) < 0\right] = \alpha. \tag{2.23}$$

Thus, the VaR corresponds to a reserve amount added to the portfolio value, such that the probability of global loss is small (equal to the level α). The VaR depends on the information available at current date t, on the conditional distributions of financial assets, on the portfolio weights, and on the loss probability level α.

It can be also viewed as an upper quantile at level $(1-\alpha)$ of the potential portfolio loss since:

$$\mathbb{P}_{S_t,t}\left[\left(V_t - V_{t+1}\right)(a_t) > VaR_{t,\alpha}(a_t)\right] = \alpha. \tag{2.24}$$

Assume that returns are normally distributed. Denote by μ_t and Σ_t its conditional mean and covariance matrix. Then, from equation (2.24), the expression of VaR is deduced:

$$VaR_{t,\alpha}(a_t) = -a_t'.\mu_t + (a_t'.\Sigma_t.a_t)^{\frac{1}{2}} q_{1-\alpha}, \tag{2.25}$$

where $q_{1-\alpha}$ is the quantile of level $(1-\alpha)$ of the standard normal distribution. Thus the VaR is the sum of conditional expected negative returns and their standard deviation, weighted by a positive constant $q_{1-\alpha}$ for a given (small) threshold α. Therefore, as seen in Chapter 3, minimizing VaR for the Gaussian case is equivalent to searching for a mean-variance efficient portfolio. $\square$

For a general position X (corresponding to a P&L), upper and lower Value-at-Risk can be defined at a given level α (see for example Acerbi [2]). Denote by F_X the cumulative distribution function (cdf) of the random variable X ($F_X(x) = \mathbb{P}[X \leq x]$). Let $q_\alpha(X)$ and $q^\alpha(X)$ be respectively the lower and upper quantiles of X.

DEFINITION 2.7 *The lower $\alpha-$ Value-at-Risk (usual VaR) is given by:*

$$VaR_\alpha(X) = -\sup\left\{x \,|F_X(x) < \alpha\right\} = -q_\alpha(X).$$

The upper $\alpha-$ Value-at-Risk is given by:

$$VaR^\alpha(X) = -\inf\left\{x \,|F_X(x) < \alpha\right\} = -q^\alpha(X). \tag{2.26}$$

REMARK 2.2 Obviously, when the cdf of X is continuous and strictly increasing, the two quantiles are equal. However, for discrete distribution functions, they may be distinct. Consider for example a P&L X with possible negative values: $-n\%$ for $-n$ in $E = \{-100, -95, ..., -5, 0\}$ with respective probabilities p_{-n}. Then, for $\alpha_n = \sum_{k=-100}^{-n} p_k$,

$$VaR_{\alpha_n}(X) = n\% \text{ and } VaR^{\alpha_n}(X) = (n+1)\%.$$

$\square$

REMARK 2.3 As mentioned in [5], note that these definitions are ambiguous. For example, a 5% lower VaR is the amount that the position may lose in the best of the 5% worst cases (and may gain in the worst of the 95% best cases). Thus, from the loss point of view, the VaR underestimates the risk at a given level α, since it provides only the smallest loss at this probability level. $\square$

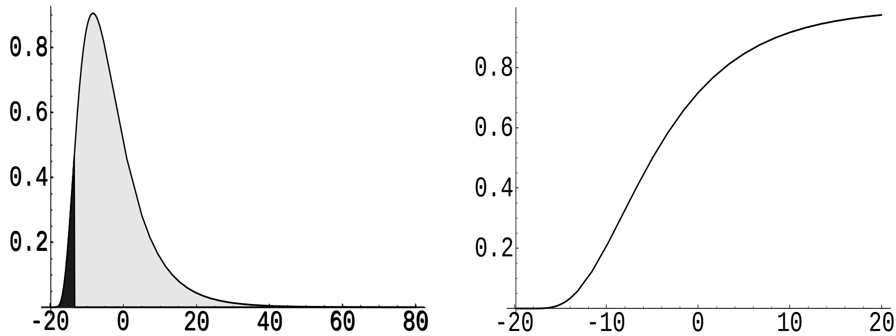


FIGURE 2.1: Probability distribution (pdf) and cumulative distribution (cdf) with $VaR_{5\%} = 13.5\%$. The 5% worst cases (resp. 95%) are shadowed in dark (resp. light) grey. In the cdf plot, VaR is the opposite of the abscissa of the intersection point of the cdf with the horizontal line $\alpha = 5\%$.

Figure (2.1) illustrates the determination of the VaR for a probability distribution having a density (pdf). For a given threshold α, the VaR_α is the opposite of the quantile q_α of the distribution: the highest ("best") value such that the probability to be under this value is smaller than α.

Figure (2.2) shows the pdf of returns and the pdf of returns conditional on exceeding the VaR. Two statistical models are considered for modelling the position: the Gaussian distribution, which has thin tails, and the Pareto-Levy distribution, whose "fat" tails decrease by a power. The two distributions have parameters such that they have the same VaR.

The Gaussian distribution is assumed to be centered and its standard deviation is chosen equal to 1%. In this case, the VaR corresponding to a 99% probability of overshoot is equal to 2.33% of the value of the position.

The Lévy-stable distribution is centered on zero and is symmetrical. Its characteristic exponent is equal to 1.5 (which determines the decrease in the distribution tail). The value of the scale parameter is equal to 1. It indicates the dispersion of the distribution. Thus the VaR of the stable Paretian distribution for a 99% probability of overshoot is also equal to 2.33%. Note that the values of the losses (lower than VaR) are near the VaR for the Gaussian distribution contrary to the stable Paretian case.

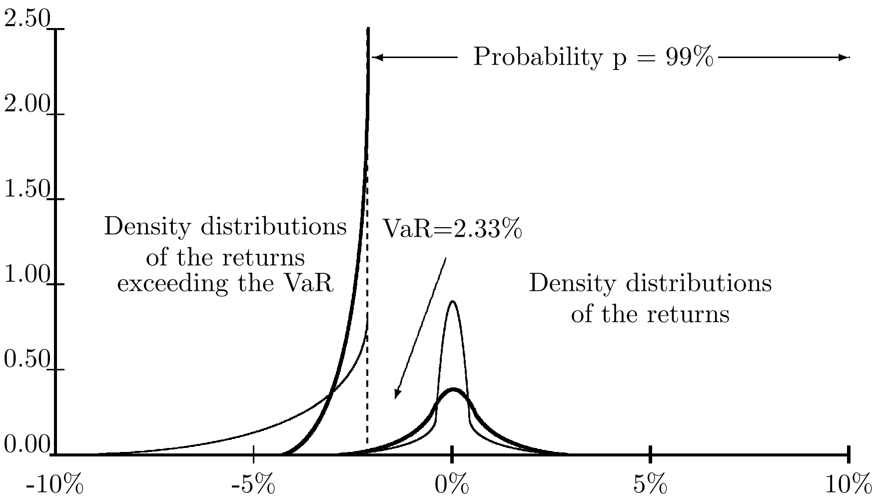


FIGURE 2.2: Level of returns for Gaussian and Stable Paretian distributions with same VaR

2.2.1.2 Convexity of the VaR

While the VaR satisfies translation invariance (T), positive homogeneity (PH), and monotonocity (M), the VaR fails to be convex for particular probability distributions. More precisely, it may not satisfy the subadditivity property, as shown in the following example provided by Danielsson et al. [144]:

Example 2.3
Consider two financial assets A and B having Gaussian probability distributions but with independent "shocks":

$$A = \epsilon + \eta,\ \epsilon \sim \mathcal{N}(\prime, \infty) \text{ and } \eta = \begin{cases} 0 & \text{with probability } 0,991, \\ -10 & \text{with probability } 0,009. \end{cases}$$

Then VaR(A) at the level 1% is equal to 3.1. Suppose that B has the same probability distribution as A but is independent from A. In that case, VaR($A + B$) at the level 1% is equal to 9.8, since for the asset $A + B$ the probability of getting -10 draw from A or B is higher than 1%.
Then we have:

$$VaR(A + B) = 9.8 > VaR(A) + VaR(B) = 3.1 + 3.1 = 6.2.$$

$\square$

Nevertheless, it may be subaddtive (and thus convex) for other probability distributions (see [30] and [144]):

Example 2.4

If the joint distribution of (X, Y) is Gaussian, then the VaR is subadditive for usual values of the level α. Indeed, as seen previously previous in Example (2.2), for a normal random variable Z, the VaR satisfies:

$$VaR_\alpha(Z) = -\left(\mathbb{E}\left[Z\right] + q_\alpha \sigma(Z)\right), \tag{2.27}$$

where q_α is the quantile of the standard Gaussian distribution and is negative since α is smaller than 0.5. Therefore, $VaR_\alpha(Z)$ is an increasing function of the standard deviation $\sigma(Z)$. Since for any (X, Y), $\sigma(X+Y) \leq \sigma(X) + \sigma(Y)$, the subaddivity property is deduced. □

2.2.1.3 Sensitivity of the VaR

The sensitivity of VaR previously in Example (2.2) has been examined in Gourieroux et al. [261], under the following assumptions: the increment of asset prices $Y_{t+1} = (S_{1,t+1} - S_{1,t}; ...; S_{N,t+1} - S_{N,t})$ has a continuous conditional distribution with positive density and finite second-order moments.

PROPOSITION 2.8

Consider the portfolio with value $V(a)$ given by $V_t(a_t) = a'.S_t$. Under previous assumptions, $VaR_{t,\alpha}(a_t)$ has a first-order derivative with respect to the allocation a given by:

$$\frac{\partial VaR_{t,\alpha}(a_t)}{\partial \alpha} = -\mathbb{E}_t\left[Y_{t+1}\left|a'Y_{t+1} = -VaR_{t,\alpha}(a_t)\right.\right]. \tag{2.28}$$

Denote by V_t the conditional variance and by $g_{a,t}$ the conditional pdf of $a'Y_{t+1}$. The second-order derivative of VaR is equal to:

$$\begin{aligned}\frac{\partial^2 VaR_{t,\alpha}(a_t)}{\partial\alpha\partial a'} &= \frac{\partial Log\left[g_{a,t}\right]}{\partial x}\left(-VaR_{t,\alpha}(a_t)\right) V_t\left[Y_{t+1}\left|a'Y_{t+1} = -VaR_{t,\alpha}(a_t)\right.\right] \\ &- \left\{\frac{\partial}{\partial x} V_t\left[Y_{t+1}\left|a'Y_{t+1} = -x\right.\right]\right\}_{x=VaR_{t,\alpha}(a_t)}.\end{aligned} \tag{2.29}$$

PROOF (see [261] for more details)

1) The VaR is defined from relation:

$$\mathbb{P}_{S_t,t}\left[A_{i,t+1} + a_{i,t}B_{i,t+1} > VaR_{t,\alpha}(a_t)\right] = \alpha,$$

where $A_{i,t+1} = -\sum_{j\neq i} a_{j,t}Y_{j,t+1}$ and $B_{i,t+1} = -Y_{i,t+1}$.
Besides, the following property holds: if (A, B) is a bivariate continuous vector, then the quantile $q(b, \alpha)$ defined by $\mathbb{P}\left[A + bB > q(b,\alpha)\right] = \alpha$, has a first-order derivative with respect to b given by:

$$\frac{\partial}{\partial b} q(b,\alpha) = \mathbb{E}\left[B\left|A + bB = q(b,\alpha)\right.\right].$$

Using this result, the first-order derivative of the VaR is deduced.

2) The second-order derivative is determined from a first-order expansion of the first-order derivative around a given allocation a_0 (see [261], Appendix B). ∎

REMARK 2.4 For the Gaussian case, the second-order derivative of the VaR is equal to

$$\frac{\partial^2 VaR_{t,\alpha}(a_t)}{\partial\alpha\partial a'} = \frac{q_{1-\alpha}}{(a_t'.\Sigma_t.a_t)^{\frac{1}{2}}} V_t\left[Y_{t+1}\left|a'Y_{t+1} = -VaR_{t,\alpha}(a_t)\right.\right],$$

where $q_{1-\alpha}$ is the quantile of level $(1-\alpha)$ of the standard normal distribution. Thus, as soon as $\alpha < 0.5$, the second-order derivative is positive which implies the convexity of the VaR. This result can be extended, for example, to a Gaussian model with unobserved heterogeneity for which the probability distribution is a mixture of Gaussian distributions (see [261]). ∎

2.2.1.4 VaR estimation

VaR estimation is obviously based on quantile estimation and tail analysis. First, parametric methods have been developed, generally using a Gaussian assumption on the joint distribution of asset returns (see *e.g.*, JP Morgan Riskmetrics). Second, to better take account of fat tails, non-parametric approaches have been introduced to determine *the historical VaR* which is an empirical quantile (see *e.g.*, [216], [283], [308]...). Semiparametric methods, based in particular on extreme value approximation for the tails have been proposed, by Bassi et al. [49] and McNeil ([380], [381]).

When observations are independent and stationary (iid), non-parametric methods based on kernel estimators can be introduced as in [261]. Consider the sequence of returns $(Z_t)_t$ defined by: for any asset i,

$$Z_{i,t} = \left(S_{i,t+1} - S_{i,t}\right)/S_{i,t}. \tag{2.30}$$

Let θ be the vector of allocations measured in values instead of shares. Then, the VaR is defined by:

$$\mathbb{P}_{S_t,t}\left[-\theta_t.Z_{t+1}) > VaR_{t,\alpha}(a_t)\right] = \alpha. \tag{2.31}$$

From iid assumption, this is equivalent to:

$$\mathbb{P}\left[-\theta_t.Z_{t+1}) > VaR_{t,\alpha}(a_t)\right] = \alpha. \tag{2.32}$$

Therefore, it can be consistently estimated from T observations by using a Gaussian kernel. The estimated VaR, denoted by $\widehat{VaR}$, is solution of:

$$\alpha = \frac{1}{T}\sum_{t=1}^{T} N\left(\frac{-\theta'.Z_t - \widehat{VaR}}{h}\right), \tag{2.33}$$

where N is the cdf of the standard normal distribution, and h is the selected bandwidth.

Equation (2.33) is solved numerically by a Gauss-Newton recursive algorithm. Let $var^{(p)}$ be the value of the approximation at step p, then:

$$var^{(p+1)} = var^{(p)} + \frac{\frac{1}{T}\sum_{t=1}^{T} N\left(\frac{-\theta'.Z_t - var^{(p)}}{h}\right)}{\frac{1}{Th}\sum_{t=1}^{T} f\left(\frac{\theta'.Z_t + var^{(p)}}{h}\right)},$$

where f is the pdf of the standard normal distribution.

The starting value $var^{(0)}$ can be chosen equal to the VaR calculated under a Gaussian assumption on the historical VaR. As mentioned in [261], this method is quite appealing and can be applied for large portfolios.

2.2.2 CVaR

As it can be easily seen, VaR is a risk measure that only takes account of the probability of losses, and not of their sizes. Moreover, VaR is usually based on the assumption of normal asset returns and has to be carefully evaluated when there are extreme price fluctuations. Furthermore, VaR may be not convex for some probability distributions. Due to these deficiencies, other risk measures have been proposed. Among them, the *Expected Shortfall (ES)* as defined in Acerbi et al. [3], also called *Conditional Value-at-Risk (CVaR) in [427]* or *TailVaR* in [31]. Note that in Acerbi and Tasche [5], several risk measures related to ES are considered and the coherence of ES is proved.

2.2.2.1 Definition of the ES

The ES can be expressed as follows (see [2]):

DEFINITION 2.8 *Let $\overleftarrow{F_X}$ be the generalized inverse of the cdf F_X of X defined by:*

$$\overleftarrow{F_X}(p) = \sup\{x\,|F_X(x) < p\}.$$

Then, the ES is defined as the average in probability of all possible outcomes of X in the probability range $0 \leq p \leq \alpha$:

$$ES_\alpha(X) = -\frac{1}{\alpha}\int_0^\alpha \overleftarrow{F_X}(p)dp. \tag{2.34}$$

Then, for continuous cdf, the ES is given by:

$$ES_\alpha(X) = -\left(\mathbb{E}\left[X\,|X \leq q_\alpha(X)\right]\right), \tag{2.35}$$

where $q_\alpha(X)$ is the quantile of X at the level α.

REMARK 2.5 Figure (2.2) illustrates how two probability distributions can have the same VaR, whereas one is thin-tailed (the Gaussian case) and the other is fat-tailed (the Pareto-Lévy case). In this example, the VaR corresponding to a 99% probability of overshoot is equal to 2.33% of the value of the position. Nevertheless, the comparison of the expected shortfalls shows the different risk levels.

As mentioned in Longin [362], for the standard Gaussian distribution, the ES is approximatively equal to $VaR + \frac{1}{2}\frac{1}{VaR}$ (here equal to 2.66%). Thus, the higher the VaR, the nearer the possible losses beyond the VaR. For a stable Paretian distribution, with location and skewness parameters equal to 0, with characteristic exponent $c > 1$, the ES is approximatively equal to $VaR + \frac{VaR}{c-1}$ (here equal to 3.22%). Therefore, the higher the VaR, the wider the spread of the losses beyond the VaR. ▯

REMARK 2.6 The expected shortfall ES is also equal to:

$$ES_\alpha(X) = -\frac{1}{\alpha}\left[\mathbb{E}\left[X 1_{\{X \le q_\alpha(X)\}}\right] - q_\alpha(X)\left(\mathbb{P}\left[X \le q_\alpha(X)\right] - \alpha\right)\right]. \quad (2.36)$$

Thus, when $F_X(x)$ has a jump at α, the term $q_\alpha(X)\left(\mathbb{P}\left[X \le q_\alpha(X)\right] - \alpha\right)$ must be substracted to the expected value $\mathbb{E}\left[X 1_{\{X \le q_\alpha(X)\}}\right]$. ▯

REMARK 2.7 When there are discontinuities in the loss distribution, the $\alpha-$tail conditional expectation TCE_α is defined by:

$$TCE_\alpha(X) = -\left(\mathbb{E}\left[X \left| X \le q_\alpha(X)\right.\right]\right), \quad (2.37)$$

is not always equal to $ES_\alpha(X)$. In that case, it may be not coherent (not subadditive) as shown in [2]. ▯

The evaluation of marginal impacts of positions on risk measures and regulatory capital is a key point for risk management analysis (see *e.g.*, example Jorion [308]).

When sensitivities are known, the global portfolio risk can be decomposed component by component. Then, the largest risk contributions can be identified. Additionally, it is no longer necessary to recompute the risk measure each time the portfolio composition is slightly modified.

The sensitivity analysis of expected shortfall and its (non-parametric) estimation has been examined in Scaillet [450]. Estimators for both expected shortfall and its first-order derivative are provided and their consistencies proved.

2.2.2.2 Sensitivity of the ES

PROPOSITION 2.9
Consider the portfolio with value $V(a)$ given by $V_t(a_t) = a'.S_t$. The expected shortfall $ES_{t,\alpha}(a_t)$ has a first-order derivative with respect to the allocation a given by:

$$\frac{\partial ES_{t,\alpha}(a_t)}{\partial \alpha} = -\mathbb{E}_t\left[Y_{t+1}\left|a'Y_{t+1} > -VaR_{t,\alpha}(a_t)\right.\right]. \qquad (2.38)$$

For the Gaussian case, the expected shortfall is given by:

$$ES_{t,\alpha}(a_t) = -a_t'.\mu_t + (a_t'.\Sigma_t.a_t)^{\frac{1}{2}}\frac{f(q_{1-\alpha})}{\alpha}, \qquad (2.39)$$

where f is the pdf of the standard Gaussian Law. Thus, its first-order derivative is equal to:

$$\frac{\partial ES_{t,\alpha}(a_t)}{\partial \alpha} = -\mu_t + \frac{\Sigma_t.a_t}{(a_t'.\Sigma_t.a_t)^{\frac{1}{2}}}\frac{f(q_{1-\alpha})}{\alpha}. \qquad (2.40)$$

PROOF (See [450]). The proof is based on the following property: consider the loss function $L_1 = X - \varepsilon Y$ where ε is a positive real number and (X, Y) is a random vector admitting a continuous conditional density with respect to the Lebesgue measure. Then the first derivative of expected shortfall is equal to:

$$\partial VaR(L_1)/\partial\varepsilon = -\mathbb{E}[Z|L_1 > VaR(L_1)]. \qquad (2.41)$$

$\square$

Note that for the Gaussian case, the first-order derivative is an affine function of the expected shortfall itself since we have:

$$\frac{\partial ES_{t,\alpha}(a_t)}{\partial \alpha} = -\mu_t + \frac{\Sigma_t.a_t}{(a_t'.\Sigma_t.a_t)}\left(ES_{t,\alpha}(a_t) + a'.\mu_t\right). \qquad (2.42)$$

2.2.2.3 Estimation of the ES

This analysis has been mainly based on estimation of the mean excess function in extreme value theory (see Embrechts et al. [201] or McNeil [380]). Thus, data have been assumed to be iid and usually univariate. In Scaillet [450], the estimation method is based on a kernel approach. This allows us to take account of more general dependencies, such as strong mixing (as defined for example in [169]).

Assume that the data have been generated by a sequence of random variables $(Y_t)_t$ which models the "risks" along the time period. The process Y is supposed to be strictly stationary. It may be, for example, a VARMA or a multivariate GARCH process.

Consider the following Kernel estimator:

$$[(Y_t, a_t'Y_t); \xi] = (Th)^{-1} \sum_{t=1}^{T} Y_t K(\xi - a_t{}'Y_t/h). \quad (2.43)$$

Introduce also:

$$\widehat{I}[\xi] = \int_{-\infty}^{\xi} [(Y_t, a_t'Y_t); u] du. \quad (2.44)$$

The bandwidth h is assumed to be a function of the number of observations T which converges to 0 as T goes to infinity. The kernel estimator $\widehat{q}(a, \alpha)$ of the quantile $-VaR_\alpha(a)$ is determined from the relation:

$$\int_{-\infty}^{\widehat{q}(a,\alpha)} [(1, a_t'Y_t); u] du = \alpha. \quad (2.45)$$

Then, the ratio $\widehat{I}[\widehat{q}(a, \alpha)]/\alpha$ is an estimator of the conditional expectation $\mathbb{E}[Y_t \,|Y_t < -VaR_\alpha(a)]$. Therefore an estimator of the ES is given by:

$$\hat{m} = -a'.\hat{I}[\hat{q}(a, \alpha)]/\alpha. \quad (2.46)$$

Similarly, an estimator of the first-order derivative of the ES is given by:

$$\hat{m}_1 = \hat{I}[\hat{q}(a, \alpha)]/\alpha. \quad (2.47)$$

Note that $\widehat{m}$ is simply equal to $\widehat{m}_1$ multiplied by the allocation a. Under mild assumptions, asymptotic normality of the estimators are proved in [450].

2.2.2.4 Numerical computation of the ES

Often, the dependence between asset prices and their particular payoffs implies that no tractable analytical expression of the portfolio distribution is available. In such case, Monte Carlo simulations can be used. Consider N independent simulations X_i of the same position value X. They allow the simulation F_X^N of the cdf X defined by:

$$F_X^N(x) = \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_i \leq x\}}. \quad (2.48)$$

Let us introduce *the ordered statistics* $(\widetilde{X}_{i,N})_i$, *i.e.*, the values $\widetilde{X}_{i,N}$ are equal to the values X_i sorted in increasing order:

$$\widetilde{X}_{1,N} \leq ... \leq \widetilde{X}_{N,N}. \quad (2.49)$$

Denote by $ES_\alpha^N(X)$ the $\alpha-$expected shortfall of the simulated distribution. For any real number x, denote by $[x]$ the integer part of x. Then:

$$ES_\alpha^N(X) = -\frac{1}{N\alpha}\left(\sum_{i=1}^{N\alpha} \widetilde{X}_{i,N}\right). \tag{2.50}$$

REMARK 2.8 Note that we have also:

$$ES_\alpha^N(X) = -\frac{1}{N\alpha}\left(\sum_{i=1}^{[N\alpha]} \widetilde{X}_{i,N} + (N\alpha - [N\alpha])\,\widetilde{X}_{[N\alpha]+1,N}\right). \tag{2.51}$$

For example, consider a sample of about 250 daily outcomes of a financial return, observed over one year. Then, with $\alpha = 5\%$, $N\alpha = 12.5$ and the $\alpha-$expected shortfall of the estimated distribution is equal to

$$ES_{5\%}^{250}(X) = -\frac{1}{12.5}\left(\sum_{i=1}^{12} \widetilde{X}_{i,250} + (0.5)\,\widetilde{X}_{13.250}\right).$$

Thus, the last term $(0.5)\,\widetilde{X}_{13,250}$ is not negligible. $\square$

Using results about convergence of ordered statistics, the following result is proved:

PROPOSITION 2.10
The $\alpha-$expected shortfall of the simulated distribution $ES_\alpha^N(X)$ converges almost surely to the $\alpha-$expected shortfall $ES_\alpha(X)$ of the true distribution. Note that when the lower and upper $\alpha-$Value-at-Risk are equal, the convergence of the $\alpha-$VaR holds: $VaR_\alpha^N(X) = -\widetilde{X}_{[N\alpha]+1,N}$ converges to $VaR_\alpha(X) = VaR^\alpha(X)$.

However, if $VaR_\alpha(X) \neq VaR^\alpha(X)$, then $VaR_\alpha^N(X)$ does not converge as N converges to infinity, but flips between $VaR_\alpha(X)$ and $VaR^\alpha(X)$.

In practice, it is necessary to also use explicit estimators of the first order derivatives of the risk measure. These derivatives are also of particular relevance in the portfolio selection problem. They help to characterize and evaluate efficient portfolio allocations when VaR and ES are substituted for variance as measures of risk. In fact, numerical constrained optimization algorithms for computations of optimal allocations usually require consistent estimates of first order derivatives in order to converge properly. Therefore, it is necessary to search for risk measures which are not only "rational," but also easily estimated and computed. This is one of the reasons to consider the *spectral risk measures.*

2.2.3 Spectral measures of risk

The expected shortfall is a "natural" coherent extension of the Value-at-Risk. Both are based on quantiles of positions X. As seen in relation (2.34), the ES is an average of the outcomes of X with respect to a particular probability which is uniform on the interval $[0, \alpha]$, *i.e.,* it assigns equal weights dp/α to the worst $100\alpha\%$ and zero to the others.

2.2.3.1 Definition of spectral risk measures

More general averages of type can be introduced as in Acerbi [1]:

$$M_\phi(X) = -\int_0^1 \phi(p)\overleftarrow{F_X}(p)dp, \tag{2.52}$$

where ϕ is a weighting function defined on the set of quantiles. In [1], the function ϕ is called *risk spectrum* or *risk aversion function,* associated to the measure M_ϕ. For the ES, the function ϕ is given by:

$$\phi_{ES_\alpha}(p) = \frac{1}{\alpha}1_{p\leq\alpha} = \begin{cases} \frac{1}{\alpha} \text{ if } p \leq \alpha, \\ 0 \text{ else.} \end{cases} \tag{2.53}$$

As any convex combination of coherent measures of risk is clearly coherent, we can search for the convex hull of a given set of such measures, among them the family of ES_α indexed by the level $\alpha \in [0, 1]$. In Acerbi [1], any measure in this convex hull is called *spectral risk measure* and the following result is proved:

PROPOSITION 2.11
Any spectral risk measure is of the type:

$$M_{\phi,c}(X) = cES_0(X) - (1-c)\int_0^1 \phi(p)\overleftarrow{F_X}(p)dp, \tag{2.54}$$

with $c \in [0, 1]$ *and* $\phi : [0, 1] \to \mathbb{R}$ *satisfying the following conditions:*
1) Positivity: $\phi(p) \geq 0$ *for all* $p \in [0, 1]$.
2) Normalization: $\int_0^1 \phi(p)dp = 1$.
3) Monotonicity: $\phi(p_1) \geq \phi(p_2)$ *for all* $0 \leq p_1 \leq p_2 \leq 1$.
The measure $M_{\phi,c}$ *is coherent if and only if these assumptions are satisfied.*

Exponential risk-aversion functions ϕ can be used:

$$\phi_\gamma(p) = \frac{e^{-p/\gamma}}{\gamma(1-e^{-p/\gamma})}, \gamma \in (0, +\infty).$$

The smaller the parameter γ, the steeper the risk-aversion function ϕ, which allows us to better take account of the left tail of the distribution.

2.2.3.2 Characterization of spectral risk measures

The set of spectral risk measures does not contain all coherent risk measures. To characterize spectral risk measures, *comonotonic additivity*, and *law-invariance* have to be introduced as in [342] and [489].

DEFINITION 2.9 *1) (Comonotonicity) Two random variables X and Y are said to be comonotonic if they are non-decreasing functions of the same random variable Z :*

$$X = f(Z) \text{ and } Y = g(Z), \text{ with } f \text{ and } g : \mathbb{R} \to \mathbb{R} \text{ non-decreasing.}$$

2) (Comonotonic additivity) A measure of risk ρ is said to be comonotonic additive if, for any comonotonic random variables X and Y:

$$\rho(X+Y) = \rho(X) + \rho(Y). \tag{2.55}$$

REMARK 2.9 The comonotonicity is a very strong dependence property between two random variables. In particular, if financial asset returns are comonotonic, they cannot hedge each other and diversification is inefficient. Thus, in that case, property (2.55) is rather intuitive. ▯

The law-invariance property is also “natural”:

DEFINITION 2.10 *A measure of risk ρ is said to be law-invariant if, for any random variables X and Y having the same probability distributions:*

$$\rho(X) = \rho(Y). \tag{2.56}$$

This means that the measure of risk ρ is a functional defined on the set of cdf:

$$\rho(X) = \rho(F_X\,[.]).$$

REMARK 2.10 As mentioned in [2], a law-invariant measure of risk gives the same value to two empirically indistinguishable random variables. Thus, a measure ρ which is not law-invariant cannot be estimated from empirical data. ▯

PROPOSITION 2.12
(see [2]) The class of spectral risk measures is the set of all coherent risk measures which are comonotonic additive and law-invariant.

This result is in favor of the spectral risk measures, since both conditions are quite rational. Basic examples of coherent measures of risk which do not satisfy one of these conditions are:

- The risk measure, introduced by Fisher [228], defined by:

$$\rho_{p,a} = -\mathbb{E}[X] + a\sigma_p^-(X), \tag{2.57}$$

  where $0 \leq a \leq 1$, and where the one-side pth moment is given by:

$$\sigma_p^-(X) = \sqrt[p]{\mathbb{E}[Max[0, (\mathbb{E}[X] - X])]^p]}. \tag{2.58}$$

  For $a \neq 0$, this measure is not comonotonic additive. Note that, for $a = 1$ and $p = 2$, the minimization of this measure is equivalent to the mean-semivariance analysis.

- The worst conditional expectation (WCE) defined in Artzner et al. [31]:

$$WCE_\alpha(X) = -\inf\{\mathbb{E}[X\,|A] : A \in \mathcal{A}, \mathbb{P}[A] > \alpha]\}, \tag{2.59}$$

  where $\mathcal{A}$ is a given σ-algebra. This measure takes the infimum of conditional risk measures only on events A with probability larger than the level α. Thus, generally it is not a law-invariant risk measure, except when the probability space is non-atomic (in that case, it is equal to the expected shortfall ES_α).

As seen in Chapter 1, the notion of stochastic dominance (in particular of the first-order) allows the comparison of two random financial positions by ordering their probability distributions. Thus, it is interesting to search for risk measures which are compatible with stochastic dominance.

DEFINITION 2.11 *A map $\rho : X \to \rho(X) \in \mathbb{R}$ is said to be monotonic with respect to first-order stochastic dominance if it satisfies:*

$$\rho(X) \geq \rho(Y) \text{ if } Y \succeq_1 X.$$

Note that if $Y \succeq_1 X$, then X always has a left tail higher than Y: $VaR_\alpha(X) \geq VaR_\alpha(Y)$ at any level α. Therefore, the previous monotonicity property is rather desirable.

Acerbi and Tasche prove that spectral risk measures are characterized by such property:

PROPOSITION 2.13
The class of spectral risk measures is the set of all coherent risk measures which are comonotonic additive and monotone with respect to first-order stochastic dominance.

2.2.3.3 Estimation of spectral risk measures

The ordered statistics $(\widetilde{X}_{i,N})_i$ allow definition of a consistent estimator $M^N_{\phi,c}(X)$ of the spectral risk measure $M_{\phi,c}(X)$, as follows: define by $ES^N_\alpha(X)$ the $\alpha-$expected shortfall of the simulated distribution by:

$$M^N_{\phi,c}(X) = -\sum_{i=1}^{N} \widetilde{X}_{i,N}\widetilde{\phi}_i \text{ with } \widetilde{\phi}_i = \int_{(i-1)/N}^{i/N} \phi(p)dp. \tag{2.60}$$

Then the sequence $\left(M^N_{\phi,c}(X)\right)_N$ converges almost surely to the spectral risk measure $M_{\phi,c}(X)$.

2.3 Further reading

Szegö ([484] and [485]) provides a general overview of risk measures and their applications to regulation, capital allocation, and portfolio optimization. Time-consistency for preferences has been introduced in Koopmans [332]. For further contributions, see Epstein and Schneider [211], Epstein and Zin [212], and Wang [502]. Examples and characterizations of time-consistent coherent risk measures are also given in Weber [503] and Cheridito *et al.* ([117], [118]), where coherent and convex monetary risk measures are defined on the space of all cadlag processes that are adapted to a given filtration, bounded or non-bounded. In Barrieu and El Karoui [43], inf-convolution of risk measures are introduced and applied to examine optimal risk transfer. When information is partial or asymmetrical with no *a priori* given probability, robust representation of convex conditional risk measures can also be deduced as in Bion-Nadal [71] (see also Detlefsen and Scandolo [165] in a dynamic framework). Law-invariant risk measures are examined in particular in Kusuoka [342]. Vector-valued coherent risk measures are introduced in Jouini *et al.* [311] to take account of difficulty in aggregating portfolio positions due to liquidity problems or transaction costs. Another approach to generate risk measures is proposed in Gooverts *et al.* [259]. This is based on insurance premium principles and uses the Markov inequality for tail probabilities. It does not always lead to coherent risk measures. This point is further analyzed by Denuit *et al.* [159], who generate a large class of risk measures from the actuarial equivalent utility pricing principle.

As examined in Wilson [504], three main methods can be used to determine the VaR:

- *The Variance/Covariance* method which assumes normal probability distribution of returns (thus with deterministic volatilities). This assumption makes easy its implementation. It is easy to interpret, to

compute, and to implement. However, it does not capture the fat tails of returns. Besides, the problem of non-linearity of derivatives payoffs has to be (partially) solved by a Delta/Gamma approximation.

- *The Historic Simulation* based on past data which are generated by "true" distributions. Fat tails and non-linearity can be taken into account. Its interpretation is straightforward, but the use of large numbers of data can be time-consuming. Moreover, the implicit assumption that past data are "good" predictors of future returns is not always valid.

- *The Monte Carlo Simulation* which has the purpose to generate values of the position from specific assumptions on returns and investment strategies. This method can be applied for quite general distributions and for different time periods. It can take non-linearity into account. However, its interpretation and computation are more involved.

Duffie and Pan [176] also provide a broad overview of VaR, in particular in the presence of market risk: asset returns have Markovian stochastic volatilities (GaRCH type, with possible regime-switching, *etc.*), and/or are modelled by jump diffusions. Factors models are also considered when the portfolio of positions has a market value which is sensitive to risk factors variations, such as major equity indices and treasury rates. A Delta/Gamma approach is also proposed to simplify the calculation methods, in particular when using Monte Carlo simulations for large portfolios.

The first derivative of the VaR and the ES with respect to portfolio allocation can be also derived when netting between positions exists, as in credit risk management (see Fermanian and Scaillet [221]).

Conditional parametric estimations of VaR and CVaR based on GARCH modelling of financial time series are examined in Alexander and Tasche [16], in McNeil *et al.* ([380], [381]), and Engle and Manganelli [207] propose a quantile estimation that does not assume normality or iid returns. They introduce a conditional autoregressive value at risk so-called *CAViaR* model where the evolution of the quantile over time is modeled by an autoregressive process.

For other detailed analyses of Value-at-Risk and its application to risk management, see Jorion [309], Dowd [170], and Stulz [482].

Part II

Standard portfolio optimization

"The rule that the investor does (or should) maximize discounted expected, or anticipated, returns...is rejected both as a hypothesis to explain, and as a maximum to guide investment behavior. We next consider the rule that the investor does (or should) consider expected return a desirable thing and variance of return an undesirable thing. This rule has many sound points, both as a maxim for, and hypothesis about, investment behavior...Diversification is both observable and sensible, a rule of behavior which does not imply the superiority of diversification must be rejected both as a hypothesis and as a maxim".

Harry Markowitz, "Portfolio Selection", *Journal of Finance*, (1952).

Most of the time, an investor (private or fund manager) has to choose a combination of securities in an uncertain framework. Two main objectives can be pursued:

- *Passive* portfolio management. The purpose is to replicate a given financial index as best as possible. The implicit assumption is that financial markets are efficient; no financial strategy can regularly outperform their performances. Clearly, for a given portfolio horizon, a private investor will choose a passive management style if the market is presumed to be bullish, with relatively high probability.

- *Active* portfolio management. If the financial market is not efficient, better asset allocations may exist. In this case, financial assets must be selected first, second their weighting must be optimized. A fund manager may also choose between two strategical approaches: the *bottom-up* process where individual securities are selected from their individual performances or the *top-down* analysis where, for example, macroeconomic and financial forecasts determine the global asset allocation among international securities.

For the first step, asset selection, the investor needs financial data and appropriate estimation methods. The selection of common stock is usually also based on earning forecasts and valuation process, such as the discounted cash flow models which allow the evaluation of, for instance, price earning ratios.

For the second step, asset weighting, the investor can select a decision rule to optimally choose the percentage invested on each asset. This choice criterion can be rationalized by using such axiomatics as detailed in Part I of this book.

However, specific constraints may be imposed on the portfolio allocation which induces more involved computational problems.

This part provides an overview of static portfolio optimization and standard performance analysis:

- First, the optimal weighting for active (but static) portfolio management, based on the seminal Markowitz model is discussed. Extensions of this theory are also presented.

- Second, indexed funds (passive management) and benchmark optimization (mixture of active and passive asset allocation) are introduced. Most of the funds are based on this latter method.

- Finally, standard performance measures to analyze and rank mutual funds are enumerated.

Chapter 3

Static optimization

This chapter is mainly devoted to the mean-variance analysis introduced by Markowitz [373]:

- The fundamental lesson of the Markowitz analysis is to show that investors must take care not only of the realized returns but also of the "risk" of their position, represented by the standard deviation of their portfolio return.

- Markowitz proposes to measure the risk of return R by its standard deviation $\sigma(R)$, and to determine the minimal $\sigma(R)$ for any fixed expected return $\mathbb{E}[R]$. The dual criterion can also be considered: maximize the expected return for any fixed standard deviation.

The first part of the chapter details the determination and the analysis of the efficient frontier:

1. Diversification property which allows for reduction of portfolio risk.
2. Optimal weights computation with the main properties of two-fund separation and efficient frontier.
3. Additional constraints on strategies, taking into account specific bounds on portfolio weights.
4. Parameter estimation problems.

The second part examines expected utility maximization and/or risk measures minimization, such as safety criteria or CVaR minimization:

1. In particular, conditions ensuring that these additional criteria allow choosing one and only one efficient portfolio are examined.
2. For expected utility maximization, results about the important property of two-fund separation are detailed.
3. For VaR/CVaR minimization, convexity properties are detailed in order to use algorithms based on convex programming.

3.1 Mean-variance analysis

Consider an investor with an initial amount V_0 to invest on given financial assets. The time horizon is fixed, the strategy is assumed to be static (*buy-and-hold*). Expectations of returns and their correlation matrix are assumed to be known.

3.1.1 Diversification effect

Example 3.1
To illustrate the impact of diversification, consider the following case: let S_1 and S_2 be two financial assets. Let R_1 and R_2 respectively be their returns. Denote by $\mathbb{E}[R_1]$ and $\mathbb{E}[R_2]$ the expectations of their returns, and by σ_1^2 and σ_2^2 their variances. Finally, their covariance is denoted by:

$$\sigma_{12} = Cov(R_1, R_2) = \rho_{12}\sigma_1\sigma_2,$$

where ρ_{12} $(-1 \leq \rho_{12} \leq +1)$ is the correlation coefficient between the two assets S_1 and S_2. Consider a portfolio P with a percentage of wealth x invested on asset S_1 and $(1-x)$ invested on S_2. Its return R_P is such that:

$$\begin{aligned} \mathbb{E}[R_P] &= x\mathbb{E}[R_1] + (1-x)\,\mathbb{E}[R_2], \\ \sigma_P^2 &= x^2\sigma_1^2 + (1-x)^2\sigma_2^2 + 2x(1-x)\,\sigma_1\sigma_2\rho_{12}. \end{aligned}$$

The following figure indicates the set of all such combinations of the two assets, according to the value of the correlation coefficient ρ_{12}. The first axis corresponds to the values of σ_P, and the second to the values of $\mathbb{E}[R_P]$.

Parameter values: $\mathbb{E}[R_1] = 10\%$, $\mathbb{E}[R_2] = 20\%$, $\sigma_1 = 0,12, \sigma_2 = 0,17$.

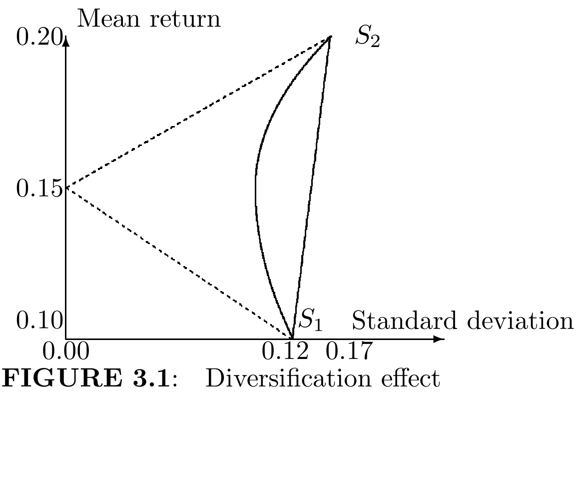


FIGURE 3.1: Diversification effect

We note that:

1) If $\rho_{12} = 1$, both expected return and standard deviation of the portfolio P are linear combinations of those of the two assets. Therefore, the set of all possible portfolios is the segment which lies between the two assets S_1 and S_2. Additionally, if $0 \leq x \leq 1$, no portfolio has a lower standard deviation than the minimal value, $\min(\sigma_1, \sigma_2)$.

2) If $-1 < \rho_{12} < +1$, we search for the minimal variance portfolio by solving the following equation:

$$\frac{\partial \sigma^2 (R_P)}{\partial x} = x\sigma_1^2 - (1-x)\sigma_2^2 + (1-2x)\sigma_1\sigma_2\rho_{12} = 0. \tag{3.1}$$

The optimal percentage is given by:

$$x^* = \frac{\sigma_2^2 - \sigma_1\sigma_2\rho_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}. \tag{3.2}$$

Its variance satisfies:

$$\sigma^2 (R_{P^*}) = \frac{\left(1-\rho_{12}^2\right)\sigma_2^2\sigma_1^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}. \tag{3.3}$$

Assume, for example, that the security S_1 is less risky than S_2 ($\sigma_1 < \sigma_2$).

Then:

$$\sigma^2 (R_{P^*}) - \sigma_1^2 = -\frac{\sigma_1^2 (\sigma_1 - \rho_{12}\sigma_2)^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}. \tag{3.4}$$

If $\rho_{12} \neq \frac{\sigma_1}{\sigma_2}$, from(3.4) we see that the difference $\sigma^2 (R_{P^*}) - \sigma_1^2$ is always negative, whatever the value of the correlation coefficient. Thus, the minimal standard deviation is smaller than σ_1.

If $\rho_{12} < \frac{\sigma_1}{\sigma_2}$, then $\sigma_2^2 - \sigma_1\sigma_2\rho_{12} > 0$ and $x^* > 0$: the percentage invested on the less risky asset is non-negative.

If $\rho_{12} > \frac{\sigma_1}{\sigma_2}$, then $\sigma_2^2 - \sigma_1\sigma_2\rho_{12} < 0$ and $x^* < 0$: this represents the short position on the less risky asset.

If $\rho_{12} = \frac{\sigma_1}{\sigma_2}$, the minimal standard deviation is equal to σ_1. In that case, there is no diversification effect.

If $\rho_{12} = -1$, there exists a portfolio with no risk: $x = \sigma_2/(\sigma_1 + \sigma_2)$.

The diversification allows the reduction of the risk as soon as the correlation coefficient between the two assets is strictly smaller than 1, except when $\rho_{12} = \frac{\sigma_1}{\sigma_2}$.

$\square$

Example 3.2
Consider a more general case with n risky securities (see Merton [387]).
Notations: for $i = 1, ..., n$,
$\mathbf{w} = (w_1, ..., w_n)$ is the vector of portfolio weights.
$\mathbf{R} = (R_1, ..., R_n)$ is the vector of asset returns.
$\overline{\mathbf{R}} = (\overline{R}_1, ..., \overline{R}_n)$ is the vector of asset returns expectations.
$\mathbf{e} = (1, ..., 1)$ is the vector with all components equal to 1.
$\mathbf{V} = [\sigma_{ij}]_{1\leq i,j\leq n}$ is the $(n \times n)$ variance-covariance matrix of returns. The matrix $\mathbf{V}$ is supposed to be invertible.

Denote by A' the vector deduced from transposition of the vector A. For each given expected return, we have to determine the minimal variance portfolio.

The expected return of any portfolio P with weights w is given by:

$$\mathbb{E}\left[R_P\right] = \sum_{i=1}^{n} w_i \mathbb{E}\left[R_i\right] = \mathbf{w}.\overline{\mathbf{R}}'. \tag{3.5}$$

The variance of the return of P is equal to:

$$\sigma^2\left(R_P\right) = \mathbf{w}'\mathbf{V}\mathbf{w} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} = 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_i w_j \sigma_{ij} + \sum_{i=1}^{n} w_i^2 \sigma_i^2. \tag{3.6}$$

The previous relation shows the decomposition of the variance of the portfolio return into two components. This relation proves that the marginal contribution of a given asset to the risk of the whole portfolio is not reduced to its own risk (its variance), but also takes account of its potential correlations to other securities. This latter property induces the diversification effect.

From relation (3.6), the partial derivative with respect to any weight w_i is deduced:

$$\frac{\partial\sigma^2\left(R_P\right)}{\partial w_i} = 2\sum_{j=1}^{n} w_j \sigma_{ij}.$$

Denote by σ_{iP} the correlation coefficient between asset i and portfolio P. Then:

$$\sum_{j=1}^{n} w_j \sigma_{ij} = \sum_{j=1}^{n} w_j Cov(R_i, R_j) = Cov(R_i, \sum_{j=1}^{n} w_j.R_j) = Cov(R_i, R_P) = \sigma_{iP},$$

and finally:

$$\frac{\partial\sigma^2\left(R_P\right)}{\partial w_i} = 2\sigma_{iP}.$$

Thus, the marginal contribution of asset i to the portfolio risk is twice its correlation with the portfolio.

Suppose now that the portfolio weights are equal to $\frac{1}{n}$. Then, its variance is given by:

$$\sigma^2(R_P) = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \sigma_i^2 + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\frac{1}{n}\right)^2 \sigma_{ij}. \tag{3.7}$$

It can be written as follows:

$$\sigma^2(R_P) = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \left[\left(\frac{1}{n}\right) \sigma_i^2\right] + \frac{2}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sigma_{ij}. \tag{3.8}$$

The first term converges to 0 when n goes to infinity (the variances $\sigma_i^2$are assumed to be uniformly upper bounded). Therefore, the individual contribution of each asset to the whole portfolio is negligible for large portfolios. If returns are independent, then the portfolio risk converges to 0; and the diversification effect is to eliminate the risk.

The second expression involves $\frac{n(n-1)}{2}$ covariances. Thus this term does not converge to 0 if the covariances σ_{ij} are also assumed to have absolute values in a given interval $[c_{\min}, c_{\max}]$ with $c_{\min} > 0$. It converges to the asymptotic mean of covariances. $\square$

3.1.2 Optimal weights

3.1.2.1 Case 1: no riskless asset

Following the Markowitz approach (see also [387]), we have to determine the set of portfolios which minimize the variance for given expected returns $\mathbb{E}[R_P]$. This leads to the following quadratic optimization problem:

$$\begin{aligned} &\min_{\mathbf{w}} \mathbf{w}'\mathbf{V}\mathbf{w}, \\ &\text{with } \mathbf{w}'\overline{\mathbf{R}} = \mathbb{E}[R_P], \\ &\mathbf{w}'\mathbf{e} = 1. \end{aligned} \tag{3.9}$$

The first constraint corresponds to the fixed expectation level. The second constraint is simply that $\mathbf{w}$ is a vector of weights. However, shortselling is allowed and no other specific constraints are introduced.

To solve Problem (3.9), consider the following Lagrangian functional:

$$L(\mathbf{w},\lambda,\delta) = \mathbf{w}'\mathbf{V}\mathbf{w} + \lambda\left(\mathbb{E}[R_P] - \mathbf{w}'\overline{\mathbf{R}}\right) + \delta(1 - \mathbf{w}'\mathbf{e}), \tag{3.10}$$

where λ and δ are the usual Lagrangian multipliers which are constant parameters. Then, Problem (3.9) is equivalent to:

$$\min_{\{\mathbf{w},\lambda,\delta\}} L(\mathbf{w},\lambda,\delta) = \mathbf{w}'\mathbf{V}\mathbf{w} + \lambda\left(\mathbb{E}[R_P] - \mathbf{w}'\overline{\mathbf{R}}\right) + \delta(1 - \mathbf{w}'\mathbf{e}). \tag{3.11}$$

The first-order conditions are:

$$\frac{\partial L(\mathbf{w},\lambda,\delta)}{\partial \mathbf{w}} = 2\mathbf{V}\mathbf{w} - \lambda\overline{\mathbf{R}} - \delta\mathbf{e} = 0, \tag{3.12}$$

$$\frac{\partial L(\mathbf{w},\lambda,\delta)}{\partial \lambda} = \mathbb{E}[R_P] - \mathbf{w}'\overline{\mathbf{R}} = 0, \tag{3.13}$$

$$\frac{\partial L(\mathbf{w},\lambda,\delta)}{\partial \delta} = 1 - \mathbf{w}'\mathbf{e} = 0. \tag{3.14}$$

Moreover, by assumption, the variance-covariance matrix $\mathbf{V}$ is invertible. Thus, the previous first-order conditions are necessary and sufficient to determine the unique solution of this linear system.

Define the four following real numbers A, B, C, and D by:

$$A = \mathbf{e}'\mathbf{V}^{-1}\overline{\mathbf{R}}, B = \overline{\mathbf{R}}'\mathbf{V}^{-1}\overline{\mathbf{R}}, C = \mathbf{e}'\mathbf{V}^{-1}\mathbf{e} \text{ and } D = BC - A^2.$$

Then, the optimal portfolio weights at the level $\mathbb{E}[R_P]$ are given by:

$$\mathbf{w} = \frac{1}{D}\left(B\mathbf{V}^{-1}\mathbf{e} - A\mathbf{V}^{-1}\overline{\mathbf{R}}\right) + \mathbb{E}[R_P]\frac{1}{D}\left(C\mathbf{V}^{-1}\overline{\mathbf{R}} - A\mathbf{V}^{-1}\mathbf{e}\right). \tag{3.15}$$

Introduce $\mathbf{w}_1$ and $\mathbf{w}_2$:

$$\mathbf{w}_1 = \frac{1}{d}\left(B\mathbf{V}^{-1}\mathbf{e} - \mathbf{A}\mathbf{V}^{-1}\overline{\mathbf{R}}\right),$$
$$\mathbf{w}_2 = \frac{1}{d}\left(C\mathbf{V}^{-1}\overline{\mathbf{R}} - A\mathbf{V}^{-1}\mathbf{e}\right).$$

Both $\mathbf{w}_1$ and $\mathbf{w}_2$ do not depend on the given level $\mathbb{E}[R_P]$. They are only determined from financial market parameters: the vector of return expectations $\overline{R}$ and the variance-covariance matrix V. Using $\mathbf{w}_1$ and $\mathbf{w}_2$, we get:

$$\mathbf{w} = \mathbf{w}_1 + \mathbb{E}[R_P].\mathbf{w}_2. \tag{3.16}$$

Therefore we deduce:

PROPOSITION 3.1

For any given return expectation level $\mathbb{E}[R_P]$, the optimal portfolio exists and is unique. Moreover, it can be decomposed as a combination of two basic portfolios, since we have:

$$\mathbf{w} = (1 - \mathbb{E}[R_P]).\mathbf{w}_1 + \mathbb{E}[R_P].(\mathbf{w}_1 + \mathbf{w}_2). \tag{3.17}$$

Any two distinct optimal portfolios generate the set of optimal portfolios.

PROOF Examine the latter assertion. Consider two optimal portfolios g and h with respective weights $\mathbf{w}_f$ and $\mathbf{w}_g$. Let q be any optimal portfolio. We have to prove that portfolio q is a combination of portfolios g and h. More precisely, we search for a real number α such that $\mathbf{w}_q = \alpha \mathbf{w}_g + (1-\alpha)\,\mathbf{w}_h$.
- Since $\mathbb{E}[R_g] \neq \mathbb{E}[R_h]$, there exists a (unique) solution α of the equation:

$$\mathbb{E}[R_q] = \alpha \mathbb{E}[R_g] + (1-\alpha)\,\mathbb{E}[R_h].$$

- The portfolio p, with weights $\{\alpha, (1-\alpha)\}$ invested on g and h, satisfies:

$$\begin{aligned}\mathbf{w}_p &= \alpha\left(\mathbf{w}_1 + \mathbf{w}_2 \mathbb{E}[R_g]\right) + (1-\alpha)\left(\mathbf{w}_1 + \mathbf{w}_2 \mathbb{E}[R_h]\right), \\ &= \mathbf{w}_1 + \mathbf{w}_2 \mathbb{E}[R_q].\end{aligned}$$

Thus $\mathbf{w}_p = \mathbf{w}_q$. ∎

REMARK 3.1 The portfolio $\mathbf{w}_1$ is associated to the expectation level $\mathbb{E}[R_P] = 1$. The portfolio $(\mathbf{w}_1 + \mathbf{w}_2)$ corresponds to the level 0. This relation proves the so-called “two mutual funds separation” property. ∎

REMARK 3.2 When the expectation level $\mathbb{E}[R_P]$ is varying, the set of optimal portfolios is a half-line included in the set of all possible portfolios. Thus, whatever the number n of securities, the optimal set is one-dimensional, whereas the set of all portfolios is $(n-1)$-dimensional (since $\sum_{i=1}^n w_i = 1$). ∎

From relation (3.15), the minimal variance at the level $\mathbb{E}(R_P)$ is computed. This yields to the fundamental implicit relation between risk and expected returns:

$$\frac{\sigma^2(R_P)}{1/C} - \frac{\left(\mathbb{E}(R_P) - A/C\right)^2}{D/C^2} = 1. \tag{3.18}$$

Geometrical interpretation.

This relation defines an arc of hyperbole in the plane with axis $(\sigma(R_P), \mathbb{E}(R_P))$.

- Its summit is the point with components $\left(\sqrt{1/C}, A/C\right)$.
- Its asymptotes are defined from the equation:

$$\mathbb{E}(R_P) = \frac{A}{C} \pm \frac{D/C^2}{1/C}\sigma(R_P).$$

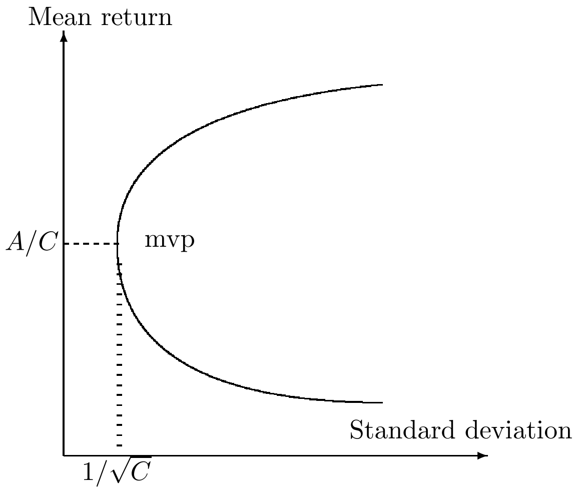


FIGURE 3.2: Mean-variance portfolios

Therefore:

- There exists one and only one portfolio with the minimal standard deviation among all possible portfolios. This corresponds to the summit of the hyperbole. It is usually called *the mean-variance portfolio* (mvp). Its expectation is equal to A/C, and its standard deviation is equal to $\sqrt{1/C}$.
- The set of all possible portfolios is delimited by the arc of the hyperbole. Thus, this one is called the portfolio frontier.
- Optimal portfolios with expected returns smaller than A/C are dominated in the mean-variance sense by those which have expected returns higher than A/C: if q is optimal and such that $\mathbb{E}\left(R_p\right) < A/C$, then there exists an optimal portfolio p with the same risk ($\sigma_{Rp} = \sigma_{R_q}$) and such that $\mathbb{E}\left(R_p\right) > A/C$.
- Consequently, for the mean-variance analysis, the "rational" portfolios are those for which the expected return is higher than the expected return of the mvp. The set of such portfolios is called *the efficient frontier*.

REMARK 3.3 (Conditions for which there exists at least an efficient portfolio with all weights that are positive.) In that case, there is no short-selling. Such results are proposed in [66], [432], and [440]. ▯

3.1.2.2 Case 2: one riskless asset

The same kind of analysis can be used when there exists a riskless asset. Denote by $\mathbf{w}$ the vector of weights of the n risky assets, and by $\mathbf{R}$ the vector of returns. The riskless asset has a return denoted by R_f. The percentage of wealth invested on this riskless asset is w_0. The budget constraint is:

$$\mathbf{w}'\mathbf{e} + w_0 = 1 \Longleftrightarrow w_0 = 1 - \mathbf{w}'\mathbf{e}. \tag{3.19}$$

Therefore, the new optimization program is:

$$\begin{gathered} \min_{\mathbf{w}} \mathbf{w}'\mathbf{V}\mathbf{w}, \\ \text{with } \mathbf{w}'\overline{\mathbf{R}} + (1 - \mathbf{w}'\mathbf{e})\, R_f = \mathbb{E}\left[R_P\right]. \end{gathered} \tag{3.20}$$

The Lagrangian functional associated to Problem (3.20) is:

$$L(\mathbf{w},\lambda) = \mathbf{w}'\mathbf{V}\mathbf{w} + \lambda\left(\mathbb{E}\left[R_P\right] - \mathbf{w}'\overline{\mathbf{R}} - (1 - \mathbf{w}'\mathbf{e})\, R_f\right). \tag{3.21}$$

Thus, we have to solve:

$$\min_{\{\mathbf{w},\lambda\}} L(\mathbf{w},\lambda). \tag{3.22}$$

The first-order conditions, which are also necessary and sufficient, are given by:

$$\frac{\partial L(\mathbf{w},\lambda)}{\partial \mathbf{w}} = 2\mathbf{V}\mathbf{w} - \lambda\left(\overline{\mathbf{R}} - \mathbf{e}R_f\right) = 0, \tag{3.23}$$

$$\frac{\partial L(\mathbf{w},\lambda)}{\partial \lambda} = \mathbb{E}\left[R_P\right] - \mathbf{w}'\overline{\mathbf{R}} - (1 - \mathbf{w}'\mathbf{e})\, R_f = 0. \tag{3.24}$$

Then the optimal portfolio at the level $\mathbb{E}\left[R_P\right]$ satisfies:

$$\mathbf{w} = \mathbf{V}^{-1}\left(\overline{\mathbf{R}} - \mathbf{e}R_f\right) \frac{\mathbb{E}\left[R_P\right] - R_f}{\left(\overline{\mathbf{R}} - \mathbf{e}R_f\right)' \mathbf{V}^{-1}\left(\overline{\mathbf{R}} - \mathbf{e}R_f\right)}. \tag{3.25}$$

Its variance is given by:

$$\sigma^2(R_P) = \mathbf{w}'\mathbf{V}\mathbf{w} = \frac{\left(\mathbb{E}\left[R_P\right] - R_f\right)^2}{J}, \tag{3.26}$$

where $J = B - 2AR_f + CR_f^2$ is non-negative.

Therefore, its standard deviation is defined as a function of the expected return level $E\left[\mathbb{R}_P\right]$:

$$\sigma(R_P) = \begin{cases} +\frac{(\mathbb{E}[R_P] - R_f)}{\sqrt{J}} & \text{if } \mathbb{E}\left[R_P\right] \geq R_f, \\ -\frac{(\mathbb{E}[R_P] - R_f)}{\sqrt{J}} & \text{if } \mathbb{E}\left[R_P\right] < R_f. \end{cases} \tag{3.27}$$

PROPOSITION 3.2

When a riskless asset is available, the optimal frontier is the union of two half-lines starting from the new mvp $(0, R_f)$, and with slopes $\sqrt{J}$ and $-\sqrt{J}$.

REMARK 3.4 The two-fund separation property shows that any efficient portfolio is a combination of the riskless asset and the tangent portfolio, which is the only efficient portfolio with nil weight on the riskless asset. The efficient frontier with the riskless asset can be decomposed into two parts:

- First, the segment which lies between theriskless portfolio and the tangential point t. This is the set of portfolios for which the weight w_0 invested on the riskless asset is positive. Investors who choose such portfolios are risk-averse: they prefer a small risk rather than a high expected return.
- Second, the half-line with origin t is the set of all efficient portfolios with a short position on the riskless asset. Investors who choose such portfolios are less risk-averse than those of the previous case: they search for higher expected returns despite higher risk.

□

Several cases can occur:

Case 1) $R_f < A/C$: the portfolio with the smallest risk (the mvp) has an expected return higher than the riskless one. This means that the financial market is favorable enough to invest on the mvp (nevertheless, its risk is not equal to 0).

PROPOSITION 3.3
In that case, the two efficient frontiers with and without the riskless asset have one intersection point t corresponding to the portfolio weight $\mathbf{w}_t$.

They are tangential at this point, defined by:

$$\mathbf{w}_t = \frac{\mathbf{V}^{-1}\left(\overline{\mathbf{R}} - \mathbf{e}R_f\right)}{(A - CR_f)}. \tag{3.28}$$

The expectation and standard deviation of this portfolio are given by:

$$\mathbb{E}\left[\mathbf{R}_t\right] = \mathbf{w}_t'\overline{\mathbf{R}}\frac{B - AR_f}{A - CR_f}, \tag{3.29}$$

$$\sigma^2\left[\mathbf{R}_t\right] = \mathbf{w}_t'\mathbf{V}\mathbf{w}_t = \frac{J}{(A - CR_f)^2}.$$

This property is illustrated by the following figure.

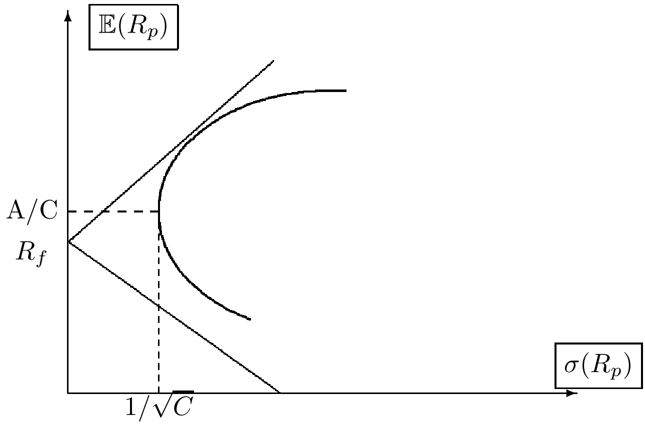


FIGURE 3.3: Efficient frontiers ($R_f < \frac{A}{C}$)

Case 2) $R_f > A/C$: the portfolio with the smallest risk (the mvp) has an expected return smaller than the riskless one. In that case, we have:

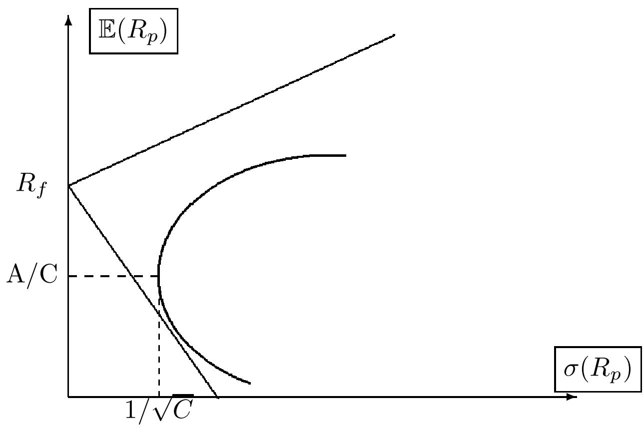


FIGURE 3.4: Efficient frontiers ($R_f > \frac{A}{C}$)

Case 3) $R_f = A/C$: Relation (3.27) gives:

$$\mathbb{E}(R_P) = \frac{A}{C} \pm \sqrt{\frac{D}{C}}\sigma(R_P). \tag{3.30}$$

This is the equation of the two asymptotes of the efficient frontier without the riskless asset. Thus, the two efficient frontiers have no intersection point.

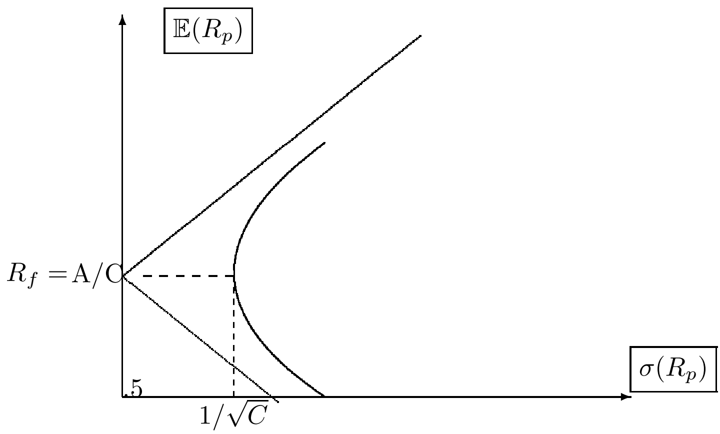


FIGURE 3.5: Efficient frontiers ($R_f = \frac{A}{C}$)

Obviously, the interesting and usual case is the first case. Nevertheless, for a given period (bearish market), case (2) can be observed.

3.1.3 Additional constraints

Transaction costs and specific constraints, such as no shortselling (see Green [265], Dybvig [181], Alexander [17]) induce a set of constraints on the investor's portfolio. When these constraints are linear, explicit solutions are still available. The optimization problem is solved by using Lagrangian methods, as shown in the previous section.

Otherwise, numerical methods must be used. Fortunately, most constraints yield to convex programming, such as the following problem:

$$\begin{gathered} \min_{\mathbf{w}} \mathbf{\Phi}(\mathbf{w}) \\ \text{with } \mathbf{w} \in \mathbf{H}, \\ \mathbf{w} \in \mathbf{K}, \end{gathered} \tag{3.31}$$

where H is a hyperplane, K is a convex set, and $\mathbf{\Phi}$ is a convex function.

Usually, H is determined from given matrix Λ and vector ν:

$$H = \{\mathbf{w} \text{ such that } \Lambda\mathbf{w} = \nu\},$$

and K is defined from a set of inequalities on convex functions:

$$K = \{\mathbf{w} \text{ such that } \Psi(\mathbf{w}) \leq \nu\} \text{ where } \Psi \text{ is a convex function.}$$

The convex optimization problems have numerical solutions[1], but the number of assets and the type of constraints may induce computational difficulties.

However, a special subclass of such constrained optimization problems are more easy to handle, *i.e.*, cone programming. Methods based on extension of the interior point algorithm can be provided, as for example in Nesterov and Nemirovski [399]. The functional Ψ is linear or quadratic, and the feasible set determined from constraints is the intersection of a hyperplane H and a cone C.

For example, the standard mean-variance problem corresponds to:

$$\Psi(\mathbf{w}) = \mathbf{w}'\mathbf{V}\mathbf{w},$$

and

$$\mathbf{\Lambda} = \begin{pmatrix} \overline{\mathbf{R}} \\ \mathbf{e} \end{pmatrix}, \nu = \begin{pmatrix} \mathbb{E}[R_P] \\ 1 \end{pmatrix}.$$

A no shortselling condition is taken into account by setting $C = \mathbb{R}^{+n}$. More general conditions, such as bounds on portfolios weights, can be introduced : $w_{\min} \leq \Gamma\mathbf{w} \leq w_{\max}$ (component by component), where $w_{\min}$ and $w_{\max}$ are fixed vectors and Γ is a given matrix. In that case, the corresponding convex K is the intersection of two cones. The choice of $w_{\min}, w_{\max}$, and Γ may depend on specific constraints, such as:

- Particular allocation on bonds and stocks, limits on international diversification, *etc.*, for institutional investors;
- Anticipated scenarios about industrial sector returns;
- In order to diversify, an upper bound of 2% can be imposed on each security; and,

[1] Special algorithms have been proposed for the mean-variance constrained optimization by Frank and Wolfe [240], and Perold [405]. See also Dantzig ([145], [146]) for linear programming, and Boyd and Vandenberghe [86]. Standard softwares propose such constrained optimization, but for relatively simple constraints.

- To limit the amount of purchase when rebalancing the portfolio between two dates t_1 and t_2, the following constraint can be imposed: let V_{t_1} the wealth at time t_1. Then the portfolio weight w_{t_2} at time t_2 satisfies:

$$V_{t_1} \times \sum_i Max(w_{t_2} - w_{t_2}, 0) \leq s,$$

where s is a fixed amount.

Example 3.3
Consider a portfolio manager who must allocate the fund among five asset classes:

1. Small caps; 2. Big caps; 3. Growth; 4. Value; and, 5. Others,

with expectations $\mathbb{E}$ and standard deviations σ as follows (percentage %):

TABLE 3.1: Expectations, variances and covariances

$\mathbb{E}$	σ	Covariances			
$\overline{R}_1 = 22$	$\sigma_1 = 21$	$\sigma_{12} = 2.8$	$\sigma_{13} = 4.00$	$\sigma_{14} = 2.30$	$\sigma_{15} = 2.70$
$\overline{R}_2 = 10$	$\sigma_2 = 14$		$\sigma_{23} = 2.60$	$\sigma_{24} = 2.00$	$\sigma_{25} = 2.10$
$\overline{R}_3 = 20$	$\sigma_3 = 20$			$\sigma_{34} = 2.05$	$\sigma_{35} = 2.75$
$\overline{R}_4 = 14$	$\sigma_4 = 12$				$\sigma_{45} = 1.70$
$\overline{R}_5 = 13$	$\sigma_5 = 15$				

- Figure 3.6 corresponds to the case where only shortselling is forbidden.
- Assume now that the investor wants to buy at least 10% of small caps, to invest at least 10% on big caps, and at most 80% on the group which contains the big caps and the values. Then, the new constrained efficient frontier is dominated by the previous one (Figure 3.7).
- Finally, if the investor wants to invest at least 10% on small caps, at least 10% on the fifth class, at least 20% on each other class, and finally, at most 80% on the group of big caps and values, then, a third constrained efficient frontier is dominated by all the previous ones (Figure 3.8).

□

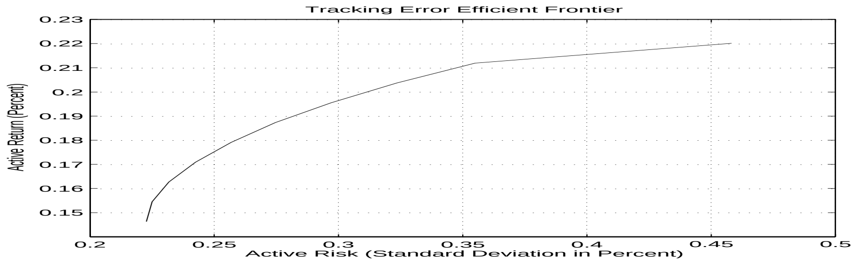


FIGURE 3.6: Efficient frontier with no shortselling

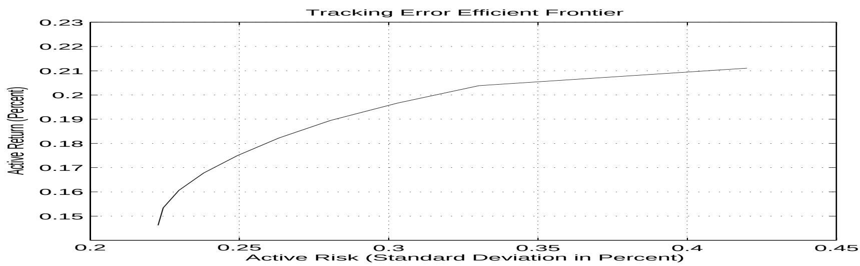


FIGURE 3.7: Efficient frontier with additional group constraints

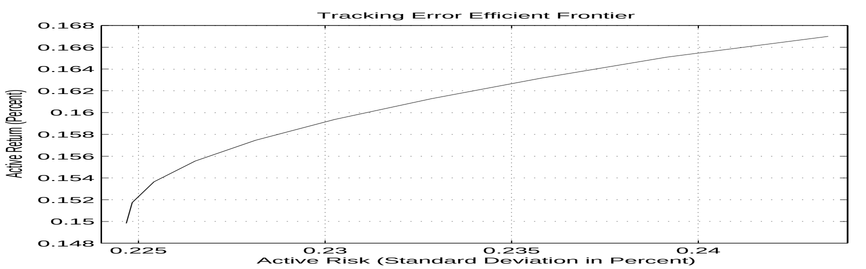


FIGURE 3.8: Efficient frontier with maximum number of constraints

3.1.4 Estimation problems

Mean-variance analysis heavily relies on "good" prediction of the mean and variance-covariance matrix of the securities. Indeed, mean-variance optimal portfolios are significantly sensitive to parameter values, as shown, for instance, in Best and Grauer [65]. These values can be estimated from financial data. However:

- Using past data to predict future returns means that we assume stability of parameters through time. Generally, the variance-covariance matrix is more stable than the mean return.

- Parameter estimations may be involved when, for example, the portfolio is composed of 150 to 250 securities.
- Numerical problems are also posed by the computation of the inverse variance-covariance matrix, as illustrated in particular by Stevens [480].

The most efficient method in terms of saving calculation time is to simplify the matrix structure. For this reason, specific hypotheses can be introduced:

- The assumption that factor models such as Sharpe's market model are valid to predict asset returns from financial indices or macroeconomics factors.
- The assumption that some correlation coefficients are identical; for example for stocks belonging to the same industrial group.

In this framework, some of the following methods can be used:

1) *Single-index and multi-index models* (see Chapter 5).

For example, assume that asset returns satisfy the Sharpe market model: for all $i \in \{1, ..., n\}$, for any $t \in [0, T]$,

$$\mathbb{E}[R_{i,t}] = \alpha_{i,t} + \beta_{i,t}\, R_{M,t} + \varepsilon_{i,t}, \tag{3.32}$$

where $R_{M,t}$ denotes the return on the market index and $\varepsilon_{i,t}$ denotes the specific risk on asset i, non-correlated with the market return. The parameters $\alpha_{i,t}$ and $\beta_{i,t}$ are determined from linear regression of the market returns on the asset returns for the same period. In particular:

$$\beta_{i,t} = Cov(R_{i,t}, R_{M,t})/Var(R_{M,t}). \tag{3.33}$$

Then the total risk of the asset i has two components: a term for systematic risk ("the market risk") and a term for non-systematic risk ("the diversifiable risk"):

$$Var(R_{i,t}) = \beta_{i,t}^2\, \sigma_{M,t}^2 + \sigma_{ej}^2 \text{ with } Var(R_{M,t}) = \sigma_{M,t}^2 \text{ and } \sigma_{ej}^2 = Var(\varepsilon_{i,t}).$$

Therefore, for a single-index model, it is sufficient to calculate the covariances of each asset with the index: for n assets, the number of operations is n instead of $n(n-1)/2$. Then, the matrix inversion is greatly simplified (see Bartlett [45]).

2) *The method of Elton, Gruber, and Padberg* ([198],[199]) (see also Chapter 9 of [200] for more details).

The method uses an optimal ranking of the assets, with the help of the simplified correlation structure. A threshold C^* is determined. Then returns below this threshold are rejected.

- The ratios $\frac{\mathbb{E}[R_i]-R_f}{\beta_i}$ are ranked from the highest to the lowest. This induces the ranking of assets: $i_1 < ... < i_n$. The higher the value of the ratio, the more desirable to include the asset in the portfolio. Therefore, if an asset is rejected, then all following assets are also rejected.

- The cut-off ratio C^* is determined through an iterative procedure, as follows. Consider the successive ratios associated to portfolios containing the l first assets. These ratios are given by:

$$C_l = \frac{\sigma_M^2 \sum_{j=1}^{l} \frac{\mathbb{E}[R_{i_j}]-R_f}{\sigma_{e_{i_j}}^2} \beta_{i_j}}{1 + \sigma_M^2 \sum_{j=1}^{l} \frac{\beta_{i_j}}{\sigma_{e_{i_j}}^2}}. \tag{3.34}$$

Note that the ratio C_l is also equal to:

$$C_l = \frac{\beta_{lM}\left(\mathbb{E}[R_M] - R_f\right)}{\beta_l}, \tag{3.35}$$

where β_{lM} is the expected change in the rate of return on asset l with 1% change in the return on the optimal portfolio.

- Consider the unique l^* for which all assets i_l such that $i_l \leq l^*$ have ratios $\left(\mathbb{E}[R_{i_l}] - R_f\right)/\beta_{i_l}$ higher than C_{i_l}, and for which all assets with $i_l > l^*$ have ratios smaller than C_{i_l}. Therefore, an asset i_l is included in the portfolio if its mean excess return, $\mathbb{E}[R_{i_l}] - R_f$, is higher than the optimal portfolio containing the first i_l assets.

3) *Average correlation models.*

The averaging data in the historical correlation matrix has been examined in [197] to forecast future values. This method assumes that the past correlation matrix provides information about what the average correlation will be in the future but contains no information about individual correlations. Therefore, the main assumption is that usual groups of industrial stocks have common correlations. Such problems can be also examined for international diversification, as in Longin and Solnik ([363],[364]).

4) *Bayesian approach.*

The Bayesian estimation process is based on one hand on prior knowledge of market parameters ("the investor's experience"), and on the other hand, on information from market observations. Then, a posterior return probability distribution can be determined. Such an approach has been examined by Stein [478], Frost and Savarino [244] and others. Black and Litterman [74] also use Bayes rule to reduce the sensitivity of the optimal allocation to parameter choices. Meucci [390] provides details about the Bayesian procedure.

REMARK 3.5 The empirical observations lead to the following conclusions (see, *e.g.*, Elton *et al.* ([198],[199]), Chan *et al.* [113], Zeng and Zhang [511].) :

- The estimation of the variance-covariance matrix is easier than the mean return estimation. Estimation from individual return data does not provide efficient forecasts. Multi-index models and average correlation models give more accurate values by reducing the impact of fluctuations (see, for example, Eun and Resnick [214]). However, significant estimation errors may remain, as mentioned by Jobson and Korkie ([305],[306]), and Chopra and Ziemba [122].

- Taking account of additional constraints, in particular the no short-selling condition, partly reduces estimation errors, as shown by Best and Grauer ([66],[67]) and by Frost and Savarino [245]. Moreover, taking more account of actual strategies of fund managers, these additional constraints often improve portfolio performances.

□

3.2 Alternative criteria

The Markowitz approach is not directly based on expected utility maximization or risk measure minimization. However, under specific assumptions, the optimal solutions of the two previous problems are mean-variance efficient portfolios.

3.2.1 Expected utility maximization

3.2.1.1 Expected utility and mean-variance analysis

The mean-variance analysis implicitly assumes that the investor's utility, defined on the portfolio return R_P, is a function $V(R_P)$ which depends only on the mean and the variance:

$$V(R_P) = f\left(\mathbb{E}\left(R_P\right), \sigma^2\left(R_P\right)\right), \tag{3.36}$$

and is increasing with respect to the mean $\left(\frac{\partial f}{\partial E} > 0\right)$, and decreasing with respect to the variance $\left(\frac{\partial f}{\partial \sigma^2} < 0\right)$. This kind of utility function is defined as a *mean-variance utility function.*

Consider an investor with an initial amount V_0 and a time horizon T who uses a "buy and hold" strategy (static portfolio optimization). In that case, the expected utility maximization is equivalent to the search of a vector of weights, $w = (w_1, ..., w_n)$, invested on n securities, which is the solution of:

$$\max_{w} \ \mathbb{E}_{\mathbb{P}}\left[U\left(V_T\right)\right]. \tag{3.37}$$

Epstein [210] proves that mean-variance utility functions are implied by a set of decreasing absolute risk aversion postulates.

For two main cases, the optimal solution is indeed mean-variance efficient:

- The utility is quadratic: $U(x) = x - \frac{k}{2}x^2$, with $k > 0$ defined on the set $]-\infty, 1/k]$ on which it is increasing. Then:

$$\mathbb{E}_{\mathbb{P}}\left[U\left(V_T\right)\right] = \mathbb{E}\left(V_0.R_P\right) - \frac{k}{2}.V_0^2.\mathbb{E}\left(R_P^2\right) = f\left(\mathbb{E}\left(R_P\right), \sigma^2\left(R_P\right)\right), \tag{3.38}$$

  with

$$f\left(\mathbb{E}\left(R_P\right), \sigma^2\left(R_P\right)\right) = U(V_0.\mathbb{E}\left(R_P\right)) - \frac{k}{2}.\sigma^2\left(R_P\right). \tag{3.39}$$

  Since the portfolio value is assumed to be in the set $]-\infty, 1/k]$, then the quantity $U(V_0.\mathbb{E}\left(R_P\right))$ is indeed an increasing function of the mean $\mathbb{E}\left(R_P\right)$. Therefore, any optimal solution is mean-variance efficient.

- The utility is exponential: $U(x) = -\frac{exp[-Ax]}{A}$ with $A > 0$, defined on $\mathbb{R}^+$. Moreover, the return distribution R is supposed to be Gaussian. Thus, the portfolio return R_P also has a Gaussian law since $R_P = \mathbf{w}.R$ and the Gaussian probability distribution is stable. Therefore, the utility $U(V_T)$ has a lognormal distribution. Recall the following property: if the random variable X has a Gaussian law $\mathcal{N}(m,\sigma)$, then

$$\mathbb{E}\left[e^X\right] = \exp\left[m + \frac{\sigma^2}{2}\right].$$

Using this expression, we deduce:

$$\mathbb{E}_{\mathbb{P}}\left[U\left(V_T\right)\right] = -(1/A)exp[-A\left(\mathbb{E}\left(R_P\right) - A/2.\sigma^2\left(R_P\right)\right)]. \tag{3.40}$$

Therefore, the maximization of this expected utility is also equivalent to the maximization of a quadratic utility with:

$$f\left(\mathbb{E}\left(R_P\right), \sigma^2\left(R_P\right)\right) = \mathbb{E}\left(R_P\right) - A/2.\sigma^2\left(R_P\right). \tag{3.41}$$

REMARK 3.6 For the previous cases, the functional to maximize has the following form:

$$V(R_P) = \mathbb{E}\left(R_P\right) - \frac{\phi}{2}.\sigma^2\left(R_P\right) \text{ where } \phi > 0. \tag{3.42}$$

The parameter ϕ is the marginal substitution rate between the mean and the variance. It can be viewed as an *aversion to the variance.* ▯

Let us examine this kind of problem. For a given aversion to variance ϕ, we have to solve:

$$\begin{array}{c} \max_{w} \ \mathbf{w}'\overline{\mathbf{R}} - \frac{\phi}{2}.\mathbf{w}'\mathbf{V}\mathbf{w}, \\ \text{with } \mathbf{w}'\mathbf{e} = 1\,. \end{array} \tag{3.43}$$

The Lagrangian associated to Problem (3.43) is defined by:

$$L\left(\mathbf{w},\lambda\right) = \mathbf{w}'\overline{\mathbf{R}} - \frac{\phi}{\mathbf{2}}.\mathbf{w}'\mathbf{V}\mathbf{w} + \lambda\left(1 - \mathbf{w}'\mathbf{e}\right) \tag{3.44}$$

The first-order conditions are given by:

$$\frac{\partial L\left(\mathbf{w},\lambda\right)}{\partial \mathbf{w}} = \overline{\mathbf{R}} - \phi\mathbf{V}\mathbf{w} - \lambda\mathbf{e} = 0, \tag{3.45}$$

$$\frac{\partial L\left(\mathbf{w},\lambda\right)}{\partial \lambda} = 1 - \mathbf{w}'\mathbf{e} = 0. \tag{3.46}$$

Solving this linear system, we deduce:

PROPOSITION 3.4
The optimal solution is equal to:

$$\mathbf{w} = \frac{1}{\phi}\left(\mathbf{V}^{-1}\overline{\mathbf{R}} - \frac{A-\phi}{C}\mathbf{V}^{-1}\mathbf{e}\right). \tag{3.47}$$

Its mean and variance are given by:

$$\mathbb{E}\left(R_P\right) = \mathbf{w}'\overline{\mathbf{R}} = \frac{d}{\phi.C} + \frac{A}{C}, \tag{3.48}$$

$$\sigma^2\left(R_P\right) = \mathbf{w}'\mathbf{V}\mathbf{w} = \frac{d}{\phi^2.C} + \frac{1}{C}. \tag{3.49}$$

When the parameter ϕ is varying in $\mathbb{R}^+$, the set of optimal solutions is exactly equal to the set of efficient portfolios.

REMARK 3.7 When the aversion to variance ϕ goes to infinity, we have:

$$\mathbb{E}\left(R_P\right) \rightarrow \frac{A}{C} \text{ and } \sigma^2\left(R_P\right) \rightarrow \frac{1}{C}. \tag{3.50}$$

In that case, the investor chooses the mvp portfolio.
When the aversion ϕ goes to 0, we have:

$$\mathbb{E}\left(R_P\right) \rightarrow +\infty \text{ and } \sigma^2\left(R_P\right) \rightarrow +\infty. \tag{3.51}$$

The smaller ϕ, the higher the mean, but also the higher the risk. ▯

REMARK 3.8 As can be seen, the main advantage of mean-variance utility functions is the simplicity of the determination of optimal solutions. Moreover, when convex constraints are added, such problems as

$$\begin{array}{c}\max\limits_{\mathbf{w}}\ \mathbf{w}'\overline{\mathbf{R}} - \frac{\phi}{2}.\mathbf{w}'\mathbf{V}\mathbf{w}, \\ \text{with } \mathbf{w}'\mathbf{e} = 1 \text{ and } \mathbf{w} \in \mathbf{K},\end{array} \tag{3.52}$$

have numerical solutions computed with efficient algorithms (see Section 3.1.3 and [405]). ▯

3.2.1.2 Optimal weights for expected utility maximization

The first-order condition of problem (3.37) is given by: for all $i \in \{1, ..., n\}$,

$$\mathbb{E}[U'\left(V_0 R_P\right)\left(R_i - R_f\right)] = 0, \tag{3.53}$$

where the portfolio return R_P is equal to $R_f + \sum_{i=1}^n w_i(R_i - R_f)$.

The solutions of the previous system of equations are generally implicit, and numerical algorithms are needed to analyze them. However, for an important class of utility functions, the HARA family, qualitative results can be derived, in particular the two-fund separation.

3.2.1.3 Two-fund separation

The mean-variance efficient portfolios have the two-fund separation property. This is not the only case for which this property is satisfied: for example, if there exists a riskless asset, other utility functions imply two-fund separation. In addition, particular probability distributions can also induce separation property.

3.2.1.3.1 Two-fund separation and utility function Indeed, Cass and Stiglitz [108] provide necessary and sufficient conditions on utility functions to get the two-fund separation.

PROPOSITION 3.5
For the static portfolio optimization problem:

1) The two-fund separation property is satisfied for any probability distribution of returns if and only if the investor has a quadratic utility function.

2) If there exists a riskless asset, then a necessary and sufficient condition for investors to have the same percentages of each risky asset is that they have an absolute risk aversion (ARA), such that for any wealth level V,

$$ARA(V) = \frac{1}{\alpha + \beta V}, \tag{3.54}$$

where the parameter β is the same for all investors and is assumed to be positive. Thus, they have the same HARA utility function (up to a linear transformation). They differ only by their wealth V_0.

PROOF We examine only the sufficient condition (see [108] for the complete proof):

Note that Condition (3.54) implies that the marginal utility U' has the form:

1) If $\beta \neq 0$, denote $\alpha = A/(\beta B)$. Then:

$$U'(V) = (A + BV)^{-\frac{1}{\beta}}.$$

2) If $\beta = 0$:

$$U'(V) = \frac{1}{\alpha} \exp\left(-\frac{V}{\alpha}\right).$$

For the HARA utility functions, note that the first-order condition (3.53) is equivalent to: if $\beta \neq 0$,

$$\mathbb{E}\left[\left(A + BV_0\left[R_f + \sum_{i=1}^{n} w_i(R_i - R_f)\right]\right)^{-\frac{1}{\beta}} (R_i - R_f)\right] = 0. \tag{3.55}$$

If $A = 0$ (CRRA case), then the optimal weights w_f and $\mathbf{w} = (w_1, ..., w_n)$ invested on the n risky assets, which are the solutions of the system of equations (3.55), do not depend on BV_0. They are the same for all investors having CRRA utility functions with the same parameter β. Denote them by:

$$w_f^{CRRA} \text{ and } w^{CRRA} = (w_1^{CRRA}, ..., w_n^{CRRA}).$$

If $A \neq 0$, define C by:

$$C = \frac{BV_0}{A + BV_0(R_f)}.$$

Then, Equation (3.55) is equivalent to:

$$\left(\frac{BV_0}{C}\right)^{-\frac{1}{\gamma}} \mathbb{E}\left[\left(1 + \sum_{i=1}^{n} Cw_i(R_i - R_f)\right)^{-\frac{1}{\gamma}} (R_i - R_f)\right] = 0.$$

Therefore, for all $i \in \{1, ..., n\}$, the value of Cw_i does not depend on the parameters of the HARA utility function.

Thus, the ratios of optimal weights, w_i^{HARA}, invested on the risky assets do not also depend on BV_0. For parameter β, they satisfy: for all $i, j \in \{1, ..., n\}$,

$$\frac{w_i^{HARA}}{w_j^{HARA}} = \frac{w_i^{CRRA}}{w_j^{CRRA}}.$$

The same result is deduced when $\beta = 0$. □

REMARK 3.9 (Determination of the utility function)
1) If $\beta \neq 0$, the utility function is given by:

$$U(V) = \frac{\beta}{(\beta - 1)B} (A + BV)^{\frac{\beta-1}{\beta}} \text{ if } \beta \neq 1,$$
$$U(V) = \frac{1}{B} \ln (A + BV) \text{ if } \beta = 1.$$

2) If $\beta = 0$:

$$U(V) = -\exp\left[-\frac{V}{\alpha}\right].$$

□

Examine the properties of optimal weights for the HARA case. Denote by $w^* = (w_1^*, ..., w_n^*)$ the percentages of risky assets in the mutual fund which depend only on the parameter β. Note that $\sum_{i=1}^{n} w_i^* = 1$. This portfolio is the optimal portfolio of an investor with a CRRA utility function ($A = 0$).

PROPOSITION 3.6

For HARA functions, the optimal weights are functions of the CRRA optimal weights.

PROPOSITION 3.7

(Determination of the optimal weights)

1) If $\beta \neq 0$, the optimal weights are given by:

$$w_i = \nu w_i^* \text{ and } w_f = 1 - \nu,$$
$$\text{with } \nu = (1-\lambda)\left(1 + \frac{A}{BV_0 R_f}\right),$$

where the parameter λ is equal to w_f for the CRRA utility function.

Thus, ν is the percentage of risky investment, and $(1-\nu)$ the percentage of riskless one.

2) If $\beta = 0$, the optimal weights are given by:

$$w_i = \left(\frac{(1-\lambda)\,\alpha}{V_0}\right) w_i^*,$$
$$w_f = \frac{\lambda\alpha}{V_0} + \frac{V_0 - \alpha}{V_0}.$$

REMARK 3.10 For the first case, note that all investors with CRRA utility functions have the same optimal portfolio (w_f and $\mathbf{w}$) whatever their initial wealth. For $A \neq 0$, the percentage ν of risky investment is decreasing w.r.t. the initial wealth. More precisely, the demand upon each risky asset i (*i.e.*, the amount invested on i) is linear in wealth. For the second case, the demand upon each risky asset i is constant. ▯

Another approach is based on the characterization of probability distributions that imply two-fund separation.

3.2.1.3.2 Two-fund separation and probability distribution Multivariate Gaussian distributions are particular solutions of this problem. They are special cases of the *spherical distributions*, defined as follows:

DEFINITION 3.1 *A random vector R has a spherical distribution if for every orthogonal map $L \in \mathbb{R}^{n\times n}$ (i.e., $L'L = LL' = I_n$ where I_n is the n-dimentional identity matrix), the random variables R and $L.R$ have the same distribution.*

This means that the distribution of a spherical random variable is invariant to rotation of the coordinates.

DEFINITION 3.2 *A random vector R is determined from its characteristic function φ_R, which must satisfy: for any vector $t \in \mathbb{R}^n$,*

$$\varphi_R(t) = \mathbb{E}\left[\exp\left(it'.R\right)\right] = \psi(t'.t),$$

where ψ is some scalar function and usually called the characteristic generator of the spherical distribution (see Fang et al. [220] for properties of such a distribution).

Examples of the spherical distributions are the Gaussian distributions, the student-t distributions, the logistic distributions, *etc.*

Chamberlain [112] provides the complete family of distributions that are necessary and sufficient for the expected utility of final wealth to be *a mean-variance utility function.*

PROPOSITION 3.8

1) If there is a riskless asset, then the distribution of any portfolio is determined by its mean and variance if and only if the random vector of returns R is a linear transformation of a spherically distributed random vector.

2) If there is no riskless asset, then the spherically distributed random vector is replaced by a random vector in which the last $n-1$ components are spherically distributed conditional on the first component, which has an arbitrary distribution. If the number of assets is infinite, then there must exist random variables X, Y, and varepsilon, where the distribution of ε conditional on X and Y is standard normal, such that every portfolio is distributed as some linear combination of X and $Y\varepsilon$. (If there is a riskless asset, then X has zero variance.)

The general result is provided by Ross [433].

PROPOSITION 3.9

For a risk-averse investor, necessary and sufficient conditions to have the two-fund separation property are the following ones:

Case 1: There is no riskless asset. The condition is that there exist two random variables $\widetilde{X}$ and $\widetilde{Y}$, and two portfolios $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$, such that the risky asset returns R_i can be written as follows: for all $i \in \{1, ..., n\}$,

$$R_i = \widetilde{X} + a_i\widetilde{Y} + \widetilde{\varepsilon}_i, \tag{3.56}$$

with:

$$\mathbb{E}\left[\widetilde{\varepsilon}_i \,\middle|\, \widetilde{X}, \widetilde{Y}\right] = 0, \quad \sum_{i=1}^{n} w_i^{(1)} \widetilde{\varepsilon}_i = \sum_{i=1}^{n} w_i^{(2)} \widetilde{\varepsilon}_i = 0, \tag{3.57}$$

$$a^{(1)} = \sum_{i=1}^{n} w_i^{(1)} a_i \neq a^{(2)} = \sum_{i=1}^{n} w_i^{(2)} a_i. \tag{3.58}$$

This condition is equivalent to the existence of two portfolios $\mathbf{w}^{(1)}$ *and* $\mathbf{w}^{(2)}$*, such that the risky asset returns* R_i *can be written as follows: for all* $i \in \{1, ..., n\}$,

$$R_i = R_i^{\mathbf{w}^{(1)}} + b_i (R_i^{\mathbf{w}^{(2)}} - R_i^{\mathbf{w}^{(1)}}) + \widetilde{\varepsilon}_i,$$

with:

$$\mathbb{E}\left[\widetilde{\varepsilon}_i \,\middle|\, R_i^{\mathbf{w}^{(2)}}, R_i^{\mathbf{w}^{(1)}}\right] = 0.$$

Case 2: There is a riskless asset with return R_f*. The conditions are the same if we set* $X = R_f$ *and* $w_f^{(1)} = 1$*,* $\mathbf{w}^{(1)} = (0, ..., 0)$*.*

PROOF We examine only the sufficient condition (see [433] for the necessary condition).

Using the risk notion of Rothschild and Stiglitz ([435],[436]) (see Chapter 1), it is sufficient to prove that any portfolio $\mathbf{w}^P$ is more risky than the combination $\mathbf{w}^{(1,2)}$ of $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$, which has the same mean return. From Condition (3.56), we have: for all $i \in \{1, ..., n\}$,

$$R_P = \sum_{i=1}^{n} w_i^P R_i = \widetilde{X} + \sum_{i=1}^{n} w_i^P a_i \widetilde{Y} + \sum_{i=1}^{n} w_i^P \widetilde{\varepsilon}_i.$$

From Condition (3.58), there exists always a real number γ^P:

$$\sum_{i=1}^{n} w_i^P a_i = \gamma^P \left(\sum_{i=1}^{n} w_i^{(1)} a_i \right) + \left(1 - \gamma^P\right) \left(\sum_{i=1}^{n} w_i^{(2)} a_i \right) = \gamma^P a^{(1)} + \left(1 - \gamma^P\right) a^{(2)}.$$

Consequently, the portfolio return R_P can be written as:

$$\begin{aligned} R_P &= \gamma^P \left(\widetilde{X} + \sum_{i=1}^{n} w_i^{(1)} a_i \widetilde{Y} \right) + \left(1 - \gamma^P\right) \left(\widetilde{X} + \sum_{i=1}^{n} w_i^{(2)} a_i \widetilde{Y} \right) + \sum_{i=1}^{n} w_i^P \widetilde{\varepsilon}_i. \\ &= \gamma^P R_i^{\mathbf{w}^{(1)}} + \left(1 - \gamma^P\right) R_i^{\mathbf{w}^{(2)}} + \sum_{i=1}^{n} w_i^P \widetilde{\varepsilon}_i. \end{aligned} \tag{3.59}$$

Consider now the portfolio $\mathbf{w}^{(1,2)} = \gamma^P \mathbf{w}^{(1)} + \left(1 - \gamma^P\right) \mathbf{w}^{(2)}$. Then:

$$R_P = R^{\mathbf{w}^{(1,2)}} + \varepsilon^P \text{ with } \varepsilon^P = \sum_{i=1}^{n} w_i^P \widetilde{\varepsilon}_i.$$

From Condition (3.57),

$$\mathbb{E}\left[\sum_{i=1}^{n} w_i^P \widetilde{\varepsilon}_i \,\middle|\, R^{\mathbf{w}^{(1,2)}}\right] = \sum_{i=1}^{n} w_i^P \mathbb{E}\left[\sum_{i=1}^{n} w_i^P \widetilde{\varepsilon}_i \,\middle|\, \widetilde{X} + \sum_{i=1}^{n} w_i^P a_i \widetilde{Y}\right] = 0.$$

Therefore, since we have:

$$R_P = R^{\mathbf{w}^{(1,2)}} + \varepsilon^P \text{ with } \mathbb{E}\left[\varepsilon^P \,\middle|\, R^{\mathbf{w}^{(1,2)}}\right] = 0,$$

the portfolio R_P is more risky than $R^{\mathbf{w}^{(1,2)}}$(according to Rothschild and Stiglitz) and cannot be optimal. Case(2) is proved in the same manner. ∎

REMARK 3.11 When the investor is risk-averse and no riskless asset is available, the two-fund separation property is based on two conditions:

- First, the return of any asset i is generated by two common factors.
- Second, there exist two portfolios, $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$, with distinct mean returns and without residual risks.
- As proved in Chamberlain [112], if for any asset i, there exists $a_i \in \mathbb{R}$ and $b_i \in \mathbb{R}$ such that:

$$R_i = a_i X + b_i Y \varepsilon \text{ with } \mathbb{E}\left[\varepsilon \,|X, Y\right] = 0, \tag{3.60}$$

then, the property of two-fund separability is satisfied, even if the utility function is not concave. ∎

3.2.2 Risk measure minimization

If investors are wary of downside risks, they want to choose portfolios with small probabilities of loss. As seen in Chapter 2, they may search to minimize the probability of having returns under a given level (this refers to VaR), or the expectation of the losses under this level (this refers to CVaR).

Attempts to solve this kind of problem are found in Roy [437], Telser [490], and Kataoka [324], when portfolio returns have Gaussian probability distributions.

3.2.2.1 Safety first

3.2.2.1.1 Roy's criterion Let $R_{\min}$ be the minimum return fixed by the investor or by statutory conditions for particular funds. Roy's criterion is the minimization of the probability to get a portfolio return lower than $R_{\min}$. Then, the optimization problem is:

$$\min_{\mathbf{w}} \mathbb{P}(R_P < R_{\min}). \tag{3.61}$$

Assume that the vector of returns has a multivariate Gaussian distribution and no riskless asset is available. Since any portfolio return also has a Gaussian

law $\mathcal{N}(\mathbb{E}[R_P], \sigma_P)$., then, by normalizing the return R_P, the Roy problem is equivalent to:

$$\min_{\mathbf{w}} \mathbb{P}\left(\frac{R_P - \mathbb{E}[R_P]}{\sigma_P} < \frac{R_{\min} - \mathbb{E}[R_P]}{\sigma_P}\right).$$

Now the probability distribution of the random variable $\dfrac{R_P - \mathbb{E}[R_P]}{\sigma_P}$ is $\mathcal{N}(0,1)$. Thus, the problem is equivalent to the minimization of $\dfrac{R_{\min} - \mathbb{E}[R_P]}{\sigma_P}$, and also to the maximization of the quantity $a = \dfrac{\mathbb{E}[R_P] - R_{\min}}{\sigma_P}$. From a geometrical point of view, in the Markowitz plane $(\sigma_P, \mathbb{E}\left[R_P\right])$, this latter ratio is the slope of a half-line starting from point $(0, R_{\min})$.

Its equation is:

$$\mathbb{E}[R_P] = a\ \sigma_P + R_{\min}.$$

Therefore, under the Gaussian assumption, the Roy portfolio is on the half-line starting from point $(0, R_{\min})$, having the maximum slope and a non-empty intersection with the set of all possible portfolios.

The solution is tangent to the efficient portfolios curve, as illustrated by the following graph:

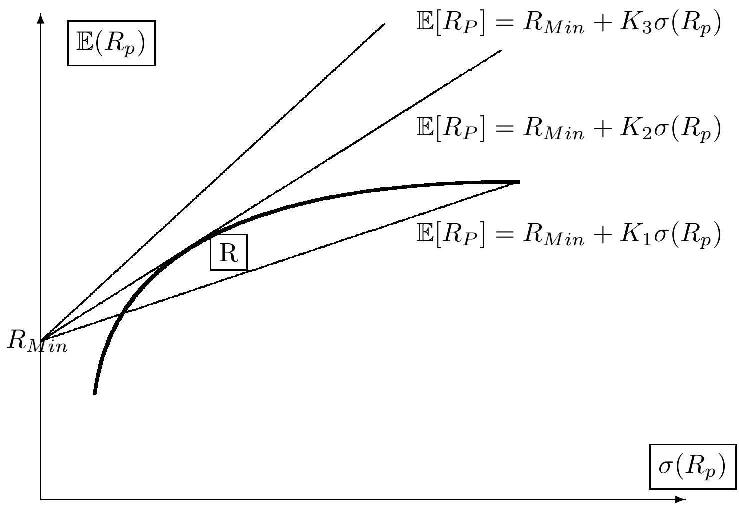


FIGURE 3.9: Roy's portfolio

The half-line with the lowest slope is mean-variance dominated by the tangent, which would be dominated by the third half-line, but this one has an empty intersection with the set of portfolios.

The determination of Roy portfolio is, for example, deduced by searching the value of the maximum ratio $a = \dfrac{\mathbb{E}[R_P] - R_{\min}}{\sigma_P}$:

1) Substituting $a\,\sigma_P + R_{\min}$ for $\mathbb{E}[R_P]$ in the hyperbole's equation induces a quadratic polynomial equation with unknown variable σ. The half-line is tangent to the efficient frontier if and only if its discriminant Δ is equal to 0.

2) Δ is itself a quadratic function of the ratio a. Thus, we have to solve this second equation. The optimal ratio a is the highest solution.

Tangent conditions can also be used to determine Roy's portfolio.

3.2.2.1.2 Telser criterion Telser's portfolio is the portfolio which has the highest return expectation under the safety conditions. We have to solve:

$$\max \mathbb{E}[R_P],$$

$$\text{with } \mathbb{P}(R_P < R_{\min}) \leq \varepsilon, \tag{3.62}$$

where both the minimum return $R_{\min}$, and the threshold ε, are fixed.

Under the Gaussian assumption, the safety condition is equivalent to:

$$\frac{R_{\min} - \mathbb{E}[R_P]}{\sigma_P} \leq x_\varepsilon,$$

where x_ε is the quantile of the standard Gaussian distribution at the level ε.

In the Markowitz plane $(\sigma_P, \mathbb{E}[R_P])$, the set of solutions is the area above the line defined by $\mathbb{E}[R_P] = R_{\min} + (-x_\varepsilon)\,\sigma_P$.

For example, if $\varepsilon = 5\%$, we have $x_\varepsilon \simeq -1.65$. Then the safety condition is approximately:

$$\mathbb{E}[R_P] \geq R_{\min} + 1.65\,\sigma_p.$$

The Telser portfolio is the intersection point of the efficient frontier with the half-line having the equation:

$$\mathbb{E}[R_P] = R_{\min} + (-x_\varepsilon)\,\sigma_P,$$

as illustrated by the following graph.

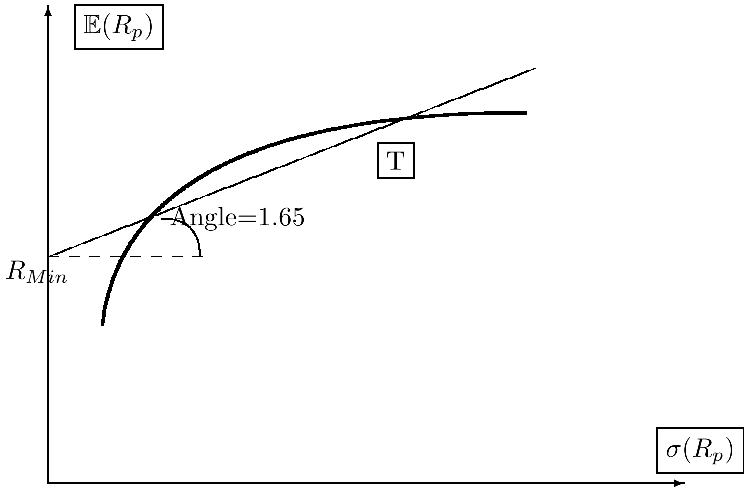


FIGURE 3.10: Telser's portfolio

3.2.2.2 Kataoka criterion

The idea of Kataoka is to search for the maximum value of the minimum return $R_{\min}$, guaranteed at a fixed probability threshold ε. We have to solve:

$$\begin{aligned} &\max_{\mathbf{w}} \; R_{\min}, \\ &\text{with } \mathbb{P}(R_P < R_{\min}) \leq \varepsilon. \end{aligned} \tag{3.63}$$

Under Gaussian assumptions, this problem is equivalent to:

$$\max_{\mathbf{w}} \; R_{\min}$$

$$\text{with } \frac{\mathbb{E}[R_P] - R_{\min}}{\sigma_P} \geq -x_{\varepsilon}.$$

In the Markowitz plane, consider the half-lines with same fixed slope $(-x_{\varepsilon})$. Then, the Kataoka portfolio is the tangent point of the efficient frontier with the unique half-line having the given slope $-x_{\varepsilon}$, and the value of $R_{\min}$ is given by the intersection with the vertical axis (see $R_{\min 2}$ in the Figure 3.11.)

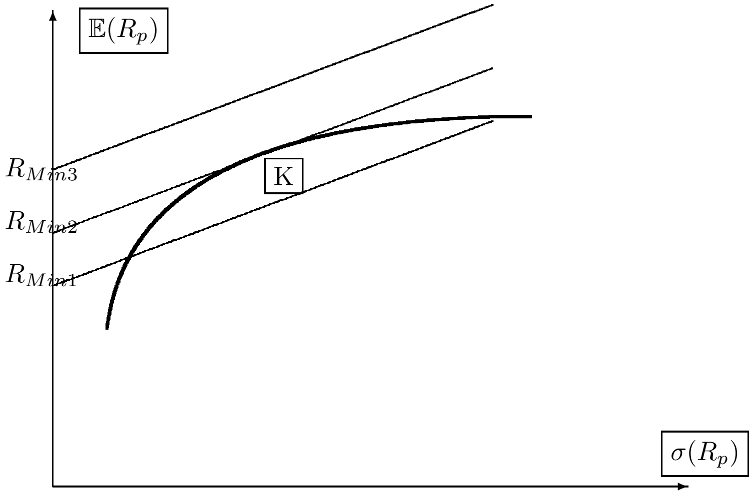


FIGURE 3.11: Kataoka's portfolio

REMARK 3.12 The three previous criteria are based on a fixed level guaranteed at a given probability threshold. Under Gaussian assumptions, all optimal solutions are necessarily mean-variance efficient. As a consequence, a unique efficient portfolio can be determined by choosing the values of $R_{\min}$ and/or the threshold ε.

- Roy criterion is based on the search of a true guarantee, since it is for this criterion that the level ε is the lowest. For example, when market volatility is high, an investor may want to reduce market risk.

- If market parameters are assumed to be well estimated, and the minimum level $R_{\min}$ and the threshold ε can be easily determined, then "rationally," the investor maximizes the return expectation. This is the purpose of Telser criterion.

- In the same framework, Kataoka portfolio has the most attractive guarantee for a given probability threshold.

□

REMARK 3.13 Obviously, other probability distributions than the Gaussian one must be used if asset returns exhibit skewness and fat tails, in particular when some payoffs are not linear with respect to given basic securities. Then, optimal solutions may no longer be mean-variance efficient. Numerical algorithms are generally needed to determine and analyze the optimal solutions, as shown in what follows. □

3.2.2.3 CVaR Minimization

Alternatively, we can search to minimize the CVaR ("expected shortfall") with or without specific constraints. For normally distributed loss functions, mean-variance (MV) and VaR/CVaR minimizations are equivalent (see computation of VaR and CVaR for the Gaussian case in Chapter 2). Otherwise, especially for asymmetrical distributions, MV and VaR/CVaR portfolio optimizations may lead to significantly different solutions. Note also that unlike in the mean-VaR optimization problem, mean-shortfall optimization can be solved efficiently as a convex optimization problem, as shown in Bertsimas *et al.* [55].

General results about VaR and CVaR minimizations are provided in Rockafellar and Uryasev ([427], [428]). They show that these problems can be based on a particular representation of the performance function, which then allows use of analytical or scenario-based optimization algorithms. When the number of scenarios is fixed, the problem is solved by linear programming or by non-smooth optimization methods.

Consider the loss function l defined on weights $\mathbf{w}$ and return R. Let us assume that the vector of returns R has a density f_R (this assumption is not critical). For $\alpha \in (0,1)$, the α-VaR denoted by $q_\alpha(\mathbf{w})$ is given by:

$$q_\alpha(\mathbf{w}) = \min \left\{ q \in \mathbb{R} : \mathbb{P}[l(\mathbf{w}, R) \geq \alpha] \right\}.$$

The α-CVaR denoted by $\phi_\alpha(\mathbf{w})$ is equal to: (notation: $x^+ = \max(x, 0)$)

$$\phi_\alpha(\mathbf{w}) = (1-\alpha)^{-1} \int_{l(w,u) \geq q_\alpha(w)} l(\mathbf{w}, u) f_R(u) du. \tag{3.64}$$

As can be seen, the α-CVaR $\phi_\alpha(\mathbf{w})$ is the conditional expectation of the loss associated to the vector of weights $\mathbf{w}$ relative to that loss being equal to $q_\alpha(\mathbf{w})$ or greater than $q_\alpha(\mathbf{w})$.

The key idea is to characterize both α-VaR $q_\alpha(\mathbf{w})$ and α-CVaR $\phi_\alpha(\mathbf{w})$ by means of a convex function $F_\alpha(\mathbf{w}, q)$ defined by:

$$F_\alpha(\mathbf{w}, q) = q + (1-\alpha)^{-1} \int_{\mathbb{R}^n} [l(\mathbf{w}, u) - q]^+ f_R(u) du. \tag{3.65}$$

We have (see [427]):

PROPOSITION 3.10

The function $F_\alpha(\mathbf{w}, q)$ is convex and continuously differentiable w.r.t. q. The α-CVaR $\phi_\alpha(\mathbf{w})$ of the loss associated with any vector of weights is given by:

$$\phi_\alpha(\mathbf{w}) = \min_{q \in \mathbb{R}} F_\alpha(\mathbf{w}, q). \tag{3.66}$$

The set $A_\alpha(\mathbf{w})$ of values q for which the minimum is attained, namely

$$A_\alpha(\mathbf{w}) = \arg\min_{q\in\mathbb{R}} F_\alpha(\mathbf{w}, q),$$

is a bounded interval which is non-empty and closed. It can be reduced to a single point. Its left endpoint is the α-VaR $q_\alpha(\mathbf{w})$. Using convexity results (see *e.g.*, Rockafellar [426]), the existence of a unique solution can be proved under specific assumptions which eliminate a local but not global minimum.

Then, a characterization of the α-CVaR can be deduced (see [427]):

PROPOSITION 3.11

The minimization of the α-CVaR $\phi_\alpha(\mathbf{w})$ is equivalent to minimizing $F_\alpha(\mathbf{w}, q)$ over all $(\mathbf{w}, q)$:

$$\min_{\mathbf{w}\in\boldsymbol{\Lambda}} \phi_\alpha(\mathbf{w}) = \min_{(\mathbf{w},q)\in\Lambda} F_\alpha(\mathbf{w}, q). \tag{3.67}$$

Moreover, any pair $(\mathbf{w}^*, q^*)$ is an optimal solution of the right-hand side if and only if $\mathbf{w}^*$ is an optimum of the left-hand side. When the interval $A_\alpha(\mathbf{w})$ reduces to a single point, any solution $(\mathbf{w}^*, q^*)$ of the minimization of the function $F_\alpha(w, q)$ is such that $\mathbf{w}^*$ minimizes the α-CVaR, and q^* is the corresponding α-VaR $q_\alpha(\mathbf{w})$.

Note also that $F_\alpha(\mathbf{w}, q)$ is convex w.r.t. $(\mathbf{w}, q)$, and $\phi_\alpha(\mathbf{w})$ is convex w.r.t. $\mathbf{w}$ if the loss function $l(\mathbf{w}, q)$ is convex w.r.t. $\mathbf{w}$. Moreover, if the set Λ of portfolios weights is convex, the optimization problem relies on convex programming (see Section 3.1.3).

REMARK 3.14 This approach allows us to avoid the VaR computation. The function $F_\alpha(w, q)$ can also be estimated from financial data or simulated from the Monte Carlo method for a given density f_R.

□

3.2.2.4 Efficient frontier

When loss functions are normally distributed, MV and VaR/CVaR optimizations generate the same efficient frontier.

Another problem is the equivalence representations of efficient frontiers with concave reward and convex risk functions.

This kind of result is known for the mean-variance case (see Steinbach [479]), and for mean-regret performance functions, see Dembo and Rosen [157].

Krokhmal et al. [336] provide a general result about this equivalence, as shown in the next proposition.

PROPOSITION 3.12
Let us consider the functions $\Phi(.)$ ("the risk") and $\Psi(.)$ ("the reward") w.r.t. the vector of weights. Consider the following three problems:

(P1) $\min_{\mathbf{w}} \Phi(\mathbf{w}) - a\Psi(\mathbf{w})$, $a \geq 0$, $\mathbf{w} \in \Lambda$,
(P2) $\min_{\mathbf{w}} \Phi(\mathbf{w})$, $\Psi(\mathbf{w}) \geq b$, $\mathbf{w} \in \Lambda$,
(P3) $\min_{\mathbf{w}} -\Psi(\mathbf{w})$, $\Phi(\mathbf{w}) \leq c$, $\mathbf{w} \in \Lambda$.

When the parameters a, b, and c are varying, three corresponding efficient frontiers are generated.

Assume that constraints $\Psi(\mathbf{w}) \geq b$ and $\Phi(\mathbf{w}) \leq c$ have internal points (under some regularity conditions from duality theory). If $\Phi(.)$ is convex, $\Psi(.)$ is concave, and Λ is convex, then the three efficient frontiers are equal. This is the case when $\Phi(.)$ is the α-CVaR $\phi_\alpha(\mathbf{w})$ and $\Psi(.)$ is the mean return.

3.3 Further reading

There is a huge amount of literature concerning static portfolio management and, in particular, mean-variance analysis, both from the theoretical and empirical points of view. The book of Elton and Gruber [200] provides a general overview about standard problems. The classic texts on portfolio optimization are the books of Markowitz ([375] and [376]). The book of Meucci [390] contains details about the Bayesian approach and computational methods when additional constraints are introduced on portfolio weights. The diversification property for mean-variance efficient portfolios is analyzed in Green and Hollifield [266]. Large-scale optimization is studied in Perold [405], Best and Kale [69], Bixby *et al.* [72], Levkovitz and Mitra [351], and Mulvey *et al.* [395]. Extension of mean-variance to more general complete markets are examined in Dybvig and Ingersoll [182]. Nowadays, mean-variance optimization is also applied on asset class level. This is due to the increasing range of asset classes, and also to the easier estimation of Markowitz inputs than for individual securities. As illustrated and mentioned for instance in Lederman and Klein [345], global asset allocation, in particular international diversification, and strategic asset allocation for long-term investment can be based on this approach.

Michaud [391] examines statistical properties and their significant impact on portfolio optimization. Ledoit and Wolf [346] examine variance-covariance matrix tests when dimensionality is large. Arch models can also give better ex-post predictions of the variance-covariance matrix. They allow for intro-

duction of factor models based on GARCH processes, which describe volatilities and correlations of securities, as shown in Herzog *et al.* [292]. Basic results about ARCH models can be found in Bollerslev *et al.* [80], Gourieroux [260].

Utility maximization is mainly developed in a dynamic framework, as can be seen in Chapters 6 and 7. Kroll *et al.* [338] examine the relation between mean-variance analysis and utility maximization. Kallberg and Ziemba [315] compare optimal portfolio weights for different utility functions. Portfolio optimization can also be based on higher-moments, involving skewness and kurtosis. A fourth-order Taylor approximation of a utility function leads to such criterion. A three-moments portfolio choice has been analyzed by Athayde and Flôres [35] for a maximum skewness portfolio. The four-moments case is examined by Jurczenko and Maillet [312], and by Malevergne and Sornette ([371],[372]). The latter articles also show how centered (absolute) moments and some cumulants are consistent measures of risks which can be used to generalize the mean-variance approach. Efficient frontiers for stochastic dominance can be also examined, as in Ruszczynski and Vanderbai [444], and Darinka and Ruszczynski [147]. Prospect theory and its connection to mean-variance analysis is examined in Lévy and Lévy [353].

When probability distributions are not Gaussian, the safety first condition can be studied by means of extreme value theory as in de Haan et al. [278]; or also by using the theory of great deviations. Giacometti and Lozza [252] examine risk measures for asset allocation. They compare, in particular, portfolio choice on both historical data and simulated returns with jointly stable non-Gaussian returns. Differences between MV and CVaR efficient frontiers are illustrated in Mausser and Rosen [379], and Anderson et al. [25] for credit risk portfolio management, and Larsen *et al.* [344], and Chabaane *et al.* [111] for hedge funds management. Gaivoronski and Pflug [251] show that VaR and CVaR efficient frontiers can be quite different. Alexander and Baptista [18] compare the mean-VaR approach with mean-variance analysis.

Other criteria can also be examined such as mean-absolute deviation (see Konno et al. ([329],[331]), minimax rule (see Cai *et al.* [101], Teo and Yang [491], and Young [508]). Athayde [34] studies the minimization of downside risk. Chekhlov *et al.* [116] provide portfolio optimization results with drawdown constraints.

Acerbi and Prospero [4] examine the minimization problem of spectral measures of risk. They prove that minimizing risks with constrained returns, or maximizing returns with constrained risks (standard risk-reward problem) coincides with the unconstrained optimization of a single suitable spectral measure. This means that "minimizing a spectral measure turns out to be already an optimization process itself, where risk minimization and returns maximization cannot be disentangled from each other."

Chapter 4

Indexed funds and benchmarking

4.1 Indexed funds

The goal of passive management is to achieve returns identical to a specific benchmark. It is based on holding a basket of securities designed, for example, to track a broad-market index which matches as closely as possible the return of the overall stock market. The most commonly used strategy is market capitalization weighting. No specific judgment is needed, and the risk of variation from the accepted asset-class benchmark is reduced as much as possible.

Why invest in global passive funds?

The first justification is the cost: Systems and personnel costs are lower than with active management since, once the procedure is in place, it is straightforward to use it. Transaction costs are also reduced. Indeed, since transactions are informationless, commissions are significantly smaller. Taxes are reduced since active portfolio strategies have a higher turnover. Moreover, bid-offer spreads are smaller because passive funds involve less small-capitalization stocks. Funds such as "trackers" also have low transaction costs.

The second is the performance: According to financial data, global indexes such as the S&P 500 and EAFE beat the median manager performance most of the time. The implicit assumption is that financial markets are efficient, and no manager has a superior performance. In fact, less than 20% of actively managed mutual funds, diversified on large-capitalizations, have outperformed the S&P 500 over the last 10 years. According to SEI Investments, more than half of the funds of the Euro area, which had performances above the average during the period 1998-2000, had lower performances than the average during the next two years.

Indexed funds are relatively recent. Until 1993, only about 5% of the total amount of mutual funds were invested on indexed funds. However, during the year 1999, about 38% of new investments have been based on passive management, which has become very popular.

When the goal is to reproduce as closely as possible a given financial index, the passive management is called an index tracking strategy. Two methods

can be considered: first to have the same assets with the same weights as the index itself (exact replication); or, second to choose a smaller subset of assets while minimizing the replication or *tracking error* for some given criterion (partial replication). For the second case, an objective function must be selected and specific constraints can be introduced, such as bounds upon individual weights, number of assets to be included in the portfolio, restrictions on transaction costs, *etc.*

4.1.1 Tracking error

Most of the time, the objective function T of the tracking error is defined as the variance of the difference between tracking portfolio return R_P and index return R_I:

$$T = \sigma(R_P - R_I). \tag{4.1}$$

Such criterion is used by Toy and Zurach [493], Connor and Leland [126], and Larsen and Resnick [344]. Other criteria, based on absolute deviations instead of squared deviations, can also be used, as in Consiglio and Zenios [127], Rockafellar and Uryasev [428], and Konno and Wijayanayake [330].

Rudolf et al. [443], for example, introduce the following "linear" *tracking-errors*:

- *MAD* (Mean Absolute Deviation).

  Let $Y \in \mathbb{R}^T$ be the index return time series. Let $X \in \mathbb{R}^{n \times T}$ be the return matrix of the n securities which are included in the replicating portfolio. Let $\beta \in \mathbb{R}^n$ the vector of weights to be determined.

  The optimal weights are deduced from the minimization of the sum of absolute deviations between index return and replicating portfolio return:

  $$\beta = Arg\min_{\beta} \sum_{t=1}^{T} \left(\left| \sum_{i=1}^{n} X_{it}\beta_i - Y_t \right| \right),$$

  where $X_t = (X_{1t}, ..., X_{nt})$.

- *MinMax.*

The optimal weights are determined against the worst case:

$$\beta = Arg\min_{\beta} \left(\max_{t} |X_t\beta - Y_t| \right).$$

These two measures of tracking-errors can be extended by selecting only the values $X_t\beta$ smaller than Y_t (*i.e.*, *Downside Risk*). This latter criterion is called MADD ("*Mean Absolute Downside Deviation*"). Similarly, we define the DMinMax criterion ("*Downside MinMax*").
To summarize, we have:

- $\mathbf{TE}_{MAD}$: $\min_{\beta} \sum_t (|X\beta - Y|)$

- $\mathbf{TE}_{MADD}$: $\min_{\beta} \sum_t \left([Y - X\beta]^+\right)$

- $\mathbf{TE}_{MinMax}$: $\min_{\beta} \max_{t} |X_t\beta - Y_t|$

- $\mathbf{TE}_{DMinMax}$: $\min_{\beta} \max_{t} [Y_t - X\beta_t]^+$

4.1.2 Simple index tracking methods

Index tracking-errors can be based on statistical methods, such as the cointegration approach, or on calibration-type algorithms, such as the threshold accepting algorithm.

4.1.2.1 Exact market capitalization weighting

This method consists of investing in all assets of the index proportional to their shares in the index. This perfect replication seems to be the ideal method. However, some difficulties may appear:

- The invested amounts must be sufficiently high in order to avoid round weighting problems.

- Depending on management style, liquidity problems can occur, for example on small-capitalizations.

- If the index is modified, high transaction costs reduce the performance.

4.1.2.2 Sratified replication

This method is based on the decomposition of the index according to a set of characteristics which can be weighted. Then, the replicating portfolio must have the same decomposition. For example, consider an equity index with $i = 1, ..., n$ stocks belonging to $j = 1, ..., m$ industrial sectors, and $k = 1, ..., l$ styles (small-cap, big-cap, growth, value, *etc.*). Then, the replicating portfolio must have the same respective percentages $(p_j\%, q_k\%)$ of sectors and styles, but not necessarily the same stocks with the same weights. However, this method does not indicate the types and the optimal number of these characteristics.

4.1.2.3 Synthetic replication

Derivatives written on the index may also be used (if they exist). However, if maturity dates do not coincide, options must be rolled over. This induces additional costs since, for example, when options are not always at-the-money, it is more difficult to forecast their prices.

4.1.2.4 Optimum sampling replication

Meade and Salkin ([382],[383]) use quadratic programming to determine the optimal tracking portfolio weights. However, they use a pre-selected set of securities. As mentioned in Tabata and Takeda [486], index fund management requires:

- Minimization of the number of assets in the tracking portfolio.
- Minimization of a function of the tracking-error between portfolio and index.

They propose an algorithm which determines a locally optimal tracking portfolio.

4.1.3 The threshold accepting algorithm

Dueck et al. ([171],[172]) have proposed the so-called *threshold accepting algorithm* for the risk and reward optimization problem. Gilli and Këllezi [255] examine the performance of the threshold accepting algorithm for index tracking. They assume the existence of transaction costs for portfolio rebalancing. Their approach is presented in this section.

4.1.3.1 The model

Assume that there are $(n_A + 1)$ securities in the index to be replicated. Asset 0 is assumed to be the risk free asset with constant rate r. Let p_{it} be the price at time t of asset i, $i = 1, ..., n_A$.

Let I_t be the index value at time t. Its return on time period $[t-1, t]$ is given by:

$$r_t^I = \ln\left(\frac{I_t}{I_{t-1}}\right).$$

Let x_{it} be the quantity invested on the ith asset in the tracking portfolio P_t at time t:

$$P_t = \{x_{it} \mid i = 0, 1, ..., n_A\}.$$

Consider the set of indices corresponding to assets included in portfolio P_t:

$$J_t = \{i \mid x_{it} \neq 0\}.$$

Denote by V_t the value of portfolio P_t:

$$V_t = \sum_{i=0}^{n_A} x_{it} p_{it} = \sum_{i \in J_t} x_{it} p_{it}.$$

The corresponding weights are given by:

$$w_{it} = \frac{x_{it} p_{it}}{V_t}.$$

Denote also by V_{t^-} the tracking portfolio value before t:

$$V_{t^-} = \sum_{i=0}^{n_A} x_{i,t-1} p_{it}.$$

Without transaction costs, the portfolio return r_t^P on time period $[t-1, t]$ is given by:

$$r_t^P = \ln\left(\frac{V_{t^-}}{V_{t-1}}\right) = \ln\left(\frac{\sum_{i=0}^{n_A} x_{i,t-1} p_{it}}{\sum_{i=0}^{n_A} x_{i,t-1} p_{i,t-1}}\right). \tag{4.2}$$

Assume that transaction costs C_t are proportional to absolute changes in each security i:

$$C_t = c \sum_{i=0}^{n_A} p_{it} \left| x_{it -} x_{i,t-1} \right|, \tag{4.3}$$

where c is a given positive real number.

Then, the portfolio return, taking account of transaction costs, is given by:

$$r_t^P = \ln\left(\frac{V_{t^-}}{V_{t-1}}\right) = \ln\left(\frac{V_{t^-}}{V_{(t-1)^-} - C_{t-1}}\right). \tag{4.4}$$

4.1.3.2 Objective function

Given the observations of prices p_{it} and I_t at times $t_1, ..., t_2$, we search for the portfolio which would have replicated as closely as possible the index return over the period $[t_1, t_2]$. Therefore:

- First we have to choose a criterion F_{t_1,t_2} which is a function of the tracking error.

- Second, we search for quantities x_{it}, $i = 1, ..., n_A$, such that the corresponding portfolio minimizes F_{t_1,t_2}. As seen previously, several measures can be used. For example, the α-norm:

$$E_{t_1,t_2} = \frac{\left(\sum_{t=t_1}^{t_2} \left| r_t^P - r_t^I \right|^{\alpha}\right)^{\frac{1}{\alpha}}}{t_2 - t_1}, \tag{4.5}$$

where $\alpha > 0$. Gilli and Këllezi [255] use this criterion for $\alpha = 1$.

4.1.3.3 Constraints

Several additional constraints can be introduced.

- No shortselling:

$$x_{it} \geqslant 0, i = 0, ..., n_A.$$

- Minimum and maximum amount on individual securities:

$$\varepsilon_i \leq \frac{x_{it} p_{it}}{\sum_{i \in J_t} x_{it} p_{it}} \leq \delta_i, i \in J_t, \tag{4.6}$$

where ε_i and δ_i are exogenous.
- Upper bound K on the number of assets:

$$\#\{J_t\} \leq K.$$

- Constraints on transaction costs: for a given non-negative coefficient γ,

$$C_t \leq \gamma V_{t^-}.$$

Other constraints can also be considered, such as limits on asset groups.

4.1.3.4 Optimization problem

Assume that at time t_2 we have the portfolio $P_{t_2^-} = \left\{x_{i,t_2^-}\right\}$ with value $V_{t_2^-}$. We attempt to "optimally" rebalance this portfolio in order to replicate the index during the time period $[t_2, t_3]$. Thus, we search for the "ideal" portfolio $P^*_{t_1} = \left\{x^*_{i,t_1}\right\}$, unchanged on $[t_1, t_2]$, which minimizes the function F_{t_1,t_2} of the tracking error on the time period $[t_1, t_2]$. Therefore, we have to solve:

$$\min F_{t_1,t_2}$$

with

$$\begin{cases} C_t \leq \gamma V_{t^-}, \\ \sum_{i \in J_t} x_{it_1} p_{it_1} + C_{t_1} = V_{t_1^-}, \\ \varepsilon_i \leq \frac{x_{i,t_1} p_{i,t_1}}{\sum_{i \in J_t} x_{i,t_1} p_{i,t_1}} \leq \delta_i, \ i \in J_t, \\ \#\{J_t\} \leq K. \end{cases}$$

Then, assuming that $P^*_{t_1}$ will be also optimal for the time period $[t_2, t_3]$, at time t_2 we choose the portfolio P_{t_2} such that $w_{i,t_2} = w^*_{i,t_1}$.

4.1.3.5 Implementation

The *TA algorithm* is a local search algorithm which avoids local extrema by accepting solutions which are not worse by more than a given threshold. This latter one is gradually decreasing and is equal to 0 after a given number of steps. The optimization problem can be defined as follows.

Let $f : \mathcal{P} \to \mathbb{R}$ be the objective function where $\mathcal{P}$ is the discrete set of all feasible replicating portfolios. Consider f_{opt} the minimum of f over the set $\mathcal{P}$:

$$f_{opt} = \min_{x \in \mathcal{P}} f(x).$$

Consider the set of all optimal solutions:

$$\mathcal{P}_{\min} = \{x \in \mathcal{P} \mid f(x) = f_{opt}\}.$$

The TA algorithm provides either a solution in $\mathcal{P}_{\min}$, or close to an element of $\mathcal{P}_{\min}$. It is based on the use of a neighborhood function, defined as follows:

DEFINITION 4.1 *-Let X be the set of all acceptable configurations for a given problem. We call "neighborhood" any function $N : X \to 2^X$, where 2^X is the class of all subsets of X.*

- A "mechanism of exploration of the neighborhood" is a procedure that indicates how we pass from a configuration $s \in X$ to a configuration $s' \in N(s)$.

- A configuration s is a local minimum with respect to the function N if $f(s) \leq f(s')$ for any configuration $s' \in N(s)$.

For the TA algorithm, at any iteration r, the acceptance of the following configuration $s' \in N(s)$ is only based on a function $r(s', s)$ and a threshold T_r; s' is accepted if $r(s', s) < T_r$.

The sequence $(T_r)_r$ is decreasing and $T_r \to 0$. We start with a given portfolio P_0. Then, the procedure indicates how to pass from one configuration, which is randomly chosen, to another which is in the neighborhood of the previous one. It takes account of the objective function. The process stops as soon as the fixed number of iterations is reached (it allows for limitation of computation time), or if another given condition is satisfied.

For example, Gilli and Këllezi [255] use the following function:

$$r(s', s) = f(s') - f(s).$$

The portfolio P_1 is determined as follows: an asset index i_1 is randomly chosen in J_{P_0} and a fixed amount of the corresponding asset is sold and converted into number of assets. Next, an asset j_1 is randomly bought. If the size constraints are satisfied by the portfolio P_0, asset j_1 is chosen in J_{P_0}. After selling i_1 and buying j_1, the portfolio constraints are examined. Adjustments are made if the constraints are not satisfied.

REMARK 4.1 The choice of assets i and j are random. However, we can search for special probability distributions in order to improve the convergence. From the theoretical point of view, results about stochastic convergence have been proved by Van Laarhoven and Aarts [497], and Zhigljavsky [512]. Beasley et al. [52] introduce an index tracking method based on genetic algorithms. A complete proof of convergence of genetic algorithms is given in Cerf [109].

$\square$

4.1.3.6 Empirical results

Gilli and Këllezi [255] test the algorithm when the exact solution is known. Using different constraints and objective functions, they prove that the TA algorithm is an efficient method of solving the index tracking and also benchmarking problems. Such a result is also shown for example in [416], both for actual financial data and for Garch simulations. To illustrate the efficiency of the TA Algorithm, consider a simulation of a multidimensional DCC-MVGARCH model (see Appendix A). The correlation matrix is estimated

from data on the period from February 2002 to May 2004. Each of the 40 stocks is modelled by a GARCH(1,1) process. The following figures illustrate the random behavior of the objective function MADD and its performance.

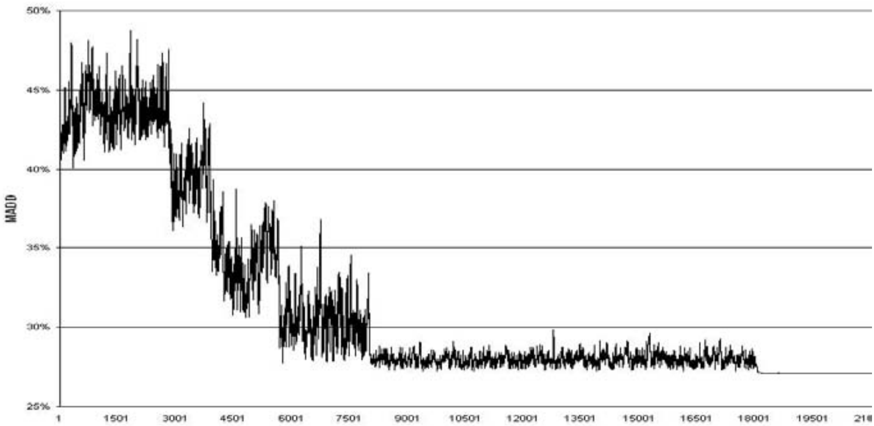


FIGURE 4.1: MADD minimization with ten stocks

On the first period, the manager applies the TA algorithm to determine the replicating portfolio weights. On the second period, the value of this portfolio is compared with the index value. The MADD minimization is quite efficient.

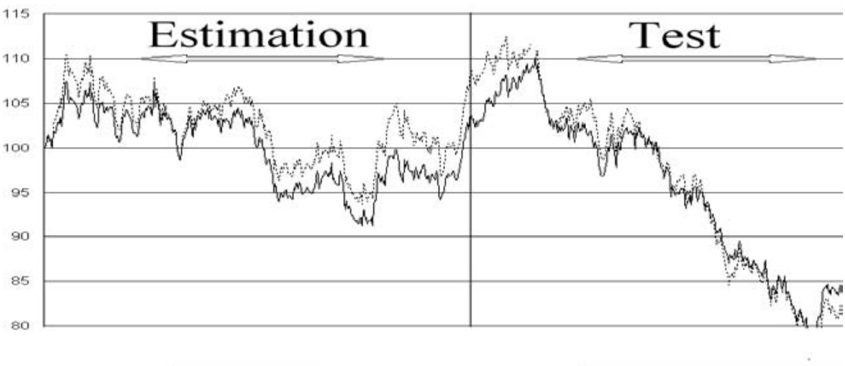


FIGURE 4.2: MADD minimization: estimation and test

However, in order to limit transaction costs, we also have to take care of the weighting stability if the manager decides to rebalance the portfolio. To illustrate this constraint, consider another DCC-MVGARCH simulation for which the manager rebalances the portfolio when one stock weight in the portfolio deviates by more than 1.5 percent (w.r.t. the corresponding weight in the index). For example, consider an initial invested amount equal to one million euros, and a proportional transaction cost of 0.5 percent. Then, we note that generally the manager will rebalance the replicating portfolio twice (rebalancing times: $T_1$and T_2,) for which the differences in the weighting are given in next table. The replicating result is quite satisfactory.

TABLE 4.1: MADD minimization and weighting differences

Stocks n^0	Initial weighting	Differences in the weighting at T_1	Differences in the weighting at T_2
1	15.45%		+0.91%
2	9.15%		+1.73%
3	15.45%		
4	13.29%	-0.64%	
5	11.52%		-1.18%
6	8.67%	+0.87%	-1.51%
7	7.05%	+0.17%	+0.42%
8	6.24%	+0.44%	+1.37%
9	6.55%	-0.21%	
10	17.63%	-1.08%	-1.84%

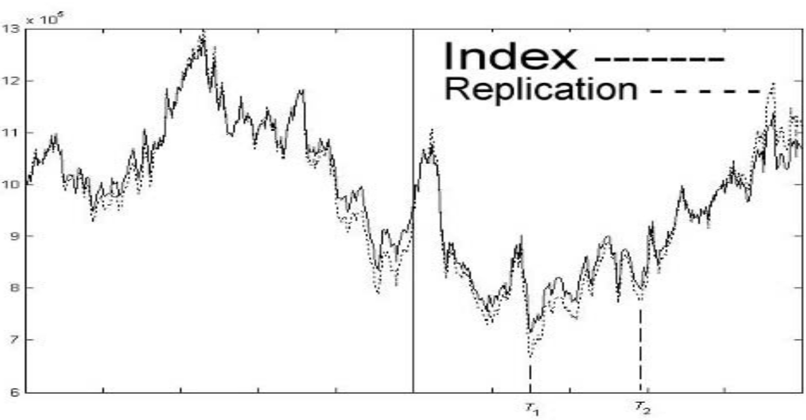


FIGURE 4.3: MADD minimization with constraints

4.1.4 Cointegration tracking method

Nowadays, the notion of cointegration is widely applied in time series analysis with applications to macroeconomics and finance. Practitioners have recently begun to use this method for indexed fund management or detection of statistical arbitrage (see Alexander and Dimitriu [15] and Dunis and Ho [179]).

Initially, the financial modelling used the analysis of return correlation coefficients. Correlation and cointegration are two related but distinct notions: correlation measures the short term dependency of two return series, while cointegration measures the long term dependency. Indeed, the correlation analysis implies dealing with stationary time series. Therefore, data series must be transformed into stationary series by deleting the trend or by differentiating them (see Granger and Joyeux [263]). Consequently, common stochastic trends between securities are ignored. The cointegration approach can use more information. Thus, this method can better describe long term relations. It is also powerful since it allows for the use of simple statistical methods such as least mean squares regressions, in order to study non-stationary processes.

In this section, some basic properties of cointegration are discussed (for basic statistical properties, see Greene [267]).

4.1.4.1 Tests of unit roots

To determine the type of non-stationarity, we have to introduce stationary tests (see Appendix A for basic statistical definitions and properties):

- Deterministic trend: function of time $(t, t^2, log\,(t)\,, ...)$. In this case, the mean is increasing (or decreasing) but the variance is constant.

- Stochastic trend: the process X is such that

$$X_t = \alpha_0 + \alpha_1.t + \varepsilon_t, \text{ with } \varepsilon_t \sim ARMA.$$

  Such process is stationary "around its mean." This stochastic trend is due to the existence of a unit root in the autoregressive process X:

$$X_t = \phi_0 + X_{t-1} + \varepsilon_t, \text{ with } \varepsilon_t \sim ARMA.$$

  Then, its variance may explode with time and its perturbations are persistent. We have to differentiate it to get a stationary time series:

$$\Delta X_t = \beta_0 + \beta_1.t + (\phi - 1)X_{t-1} + \varepsilon_t, \text{with } \varepsilon_t \sim WN(0, \sigma^2).$$

  The stationary test (called the Dickey-Fuller test) is the following:

- Null hypothesis: $(H_0) : \phi - 1 = 0,$

- Alternative hypothesis: $(H_1) : \phi - 1 < 1$.

If assumption H_1 is rejected, then the process is considered non-stationary. Coefficients β_1 and β_0 allow us to check if there exists a deterministic/stochastic trend. A special statistic has been introduced by Dickey and Fuller [180] (see also Davidson and Mac-Kinnon [178]). However, this test is limited since it assumes that the noise is a white noise. Thus, we have to use the "Augmented Dickey-Fuller" (ADF) test.

Additional lagged variables are introduced:

$$\Delta X_t = \beta_0 + \beta_1.t + (\phi - 1)X_{t-1} + \sum_{j=1}^{p} \lambda_j.\Delta X_{t-j} + \varepsilon_t, \text{ with } \varepsilon_t \sim WN(0, \sigma^2).$$

The number of lags p for the variable ΔY_{t-j} is chosen such that ε is a white noise.

REMARK 4.2 For a stationary time series, we have $\phi = 0$. When d roots exist, we call $I(d)$ the time series X. It must be differentiated d times in order to get a stationary series (see Granger [264]). Note that financial time series are generally such that $d \leq 2$.

□

4.1.4.2 Cointegration definition

Consider two time series $(X_t)_t$ and $(Y_t)_t$. If they are both $I(1)$, any linear combination of these two variables is generally not stationary. However, in some cases, this combination is $I(0)$, which means that it is stationary. In that case, the two series $(X_t)_t$ and $(Y_t)_t$ are said to be *cointegrated*:

$$Y_t = a + b.X_t + Z_t \text{ where } Y_t \sim I(1),\ X_t \sim I(1) \text{ and } Z_t \text{ is stationary.} \quad (4.7)$$

This means that these two series have similar trends such that, for a specific linear combination, their trends can be compensated to provide a stationary time series. If Z_t is strongly stationary, then the difference $Y_t - (a + b.X_t)$ has always the same probability distribution. There exists a "long term equilibrium" between the two series.

From relation (4.7), to test the cointegration property, it is sufficient to test the stationarity of the residual terms z_t.

This can be done by using Dickey-Fuller tests such as:

$$\Delta z_t = (\varphi - 1)z_{t-1} + \nu_t.$$

4.1.4.3 Tracking portfolio determination

This problem is solved by using the following three steps:

- *First step.* We have to select a given subset of assets, either by means of, for example, "stock picking,"or by statistical methods. This step is essential to get a significant property of cointegration, which determines the quality of the tracking method. It cannot be based on cointegration. The manager has to test different securities combinations to be included in the replicating portfolio. In particular, he must select the number of assets. The higher this number, the more stable the cointegration (in the absence of costs).

- *Second step.* The logarithm of the index value is regressed by least squares estimation (LSE) on the logarithms of the previously selected asset prices:

$$\log(I_t) = c_1 + \sum_{k=1}^{n} c_{k+1} \log(P_{k,t}) + \varepsilon_t. \tag{4.8}$$

  Note that by the logarithm transformation, time series are more homogeneous. If time series are cointegrated, then their logarithms are also cointegrated. The residuals are stationary if and only if $\log(I)$ and the tracking portfolio $\sum_{k=1}^{n} c_k. \log(P_k)$ are cointegrated. The coefficients c_k are the weights of the tracking portfolio. In order to have a cointegration relation, the residual term ε in relation (4.8) must be $I(0)$ and all the asset prices in the tracking portfolio are $I(d)$ with $d \geqslant 1$. Otherwise, the coefficients c_k, determined from LSE, are not efficient and may be fallacious.

- *Third step.* We have to estimate the following dynamic of index return:

$$\Delta \log(I_t) = \gamma \widehat{\varepsilon}_{t-1} + \sum_{h=1}^{n} \alpha_h \Delta \log(I_{t-h}) + \sum_{k \in CI} \Gamma_k \Delta \log(P_{k,t}) + u_t. \tag{4.9}$$

  This king of regression is an "error correction model": it corrects short term effects.

  The coefficient γ models the speed of mean-reverting to the long term value, given by the cointegration relation. It be must non-positive. Coefficients Γ_k are normalized such that $\sum_{k=1}^{n} \Gamma_k = 1$, since they represent the weights of the tracking portfolio.

4.1.4.3.1 First step: causality To determine the assets to introduce in the tracking portfolio, the notion of causality between the tracking portfolio and the index can be introduced. This notion has been considered in Granger [262]. A series induces causality on one another "if the knowledge of the first one improves the forecast of the second one."

Granger causality test. This is based on linear regressions.

DEFINITION 4.2 *Let X be a multivariate process $X_t = (X_{1,t}, ...X_{n,t})'$. The random variable X_j does not cause X_k if and only if, at any time t, the knowledge of the past of X_j, $\underline{X_{j,t-1}}$, does not improve the forecast of $X_{k,t+H}$, for any time horizon H:*

$$\mathbb{EL}\left(X_{k,t+H}/X_{t-1}; 1 \leq i \leq n\right) = \mathbb{EL}\left(X_{k,t+H}/\underline{X_{j,t-1}}\right), \tag{4.10}$$

where $\mathbb{EL}(./.)$ is the linear regression operator, and $\underline{X_{j,t-1}}$ is the set of random variables $\{X_{i,t-1}; i \neq j\}$.

In practice, for any asset i, consider the following regression on the rates of return:

$$r_t^I = \alpha_0 + \sum_{p=1}^{k} \alpha_p r_{t-p}^I + \sum_{p=1}^{k} \beta_p r_{t-p}^i + \epsilon_t. \tag{4.11}$$

To determine the optimal value of the number k of assets, the Akaike criterion (AIC) [11] can be used:

$$AIC = -2(\frac{L}{T}) + 2\frac{k}{T}, \tag{4.12}$$

where L is the likelihood, and T is the number of observations. The Granger test on Equation (4.11) for any asset i, are based on a Fisher test on the joint hypothesis:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0. \tag{4.13}$$

Sims causality Another definition of causality is given by Sims [471]. It is based on impulse analysis. A time series causes another one if a random shock on the first one has an impact on the other one, especially on the variance of its forecast errors about its future values. To characterize the Sims causality between two stationary time series, consider the Wold canonical decomposition:

$$X_t = (X_{1t}, ..., X_{nt})' = C(L)\varepsilon_t = C(L)(\varepsilon_{1t}, ..., \varepsilon_{nt})', \tag{4.14}$$

where L is the lag operator, $C(0) = I_d$, and ε_t is the canonical innovation of X_t which has the variance-covariance matrix $\Sigma = (\sigma_{ij})_{1 \leq i,j \leq n}$:

$$\varepsilon_{it} = X_{it} - \mathbb{EL}\left(X_{it}/\underline{X_{i,t-1}}\right). \tag{4.15}$$

The random variable X_t can be decomposed on two moving averages, according to:

$$X_{it} = \sum_{j=1}^{n} C_{ij}(L)\,\varepsilon_{jt}. \tag{4.16}$$

According to Sims, at any time t, past shocks on variables X_j have an impact on X_{it}, through the variable $C_{ij}(L)\varepsilon_{jt}$. This approach allows us to measure the influence of different impulses on variables. Therefore, the manager can select securities to be included in the tracking portfolio by using two causality criteria: the first one based on forecast improvement, the second on impulse reaction.

4.1.4.3.2 Second step: cointegration tests The first method, due to Engle and Granger [205], uses the following approach. We begin to search for a long term relation between a dependent variable and explicative variables. Then we have to introduce an error correcting model for the cointegrated variables. Nevertheless, this method assumes prior relations and special causality between the variables.

The second method uses the maximum likelihood, introduced by Johansen [304]. Since this method uses a multivariate model, it allows for differentiating of several cointegrated vectors. Stock and Watson [481] prove that under the cointegration hypothesis, (*i.e.*, ε is stationary), the estimators of the coefficients c_i in Equation (4.8) have a very fast speed of convergence: it is equal to T and not to $\sqrt{T}$, as is usual.

4.1.4.3.3 Third step: stability tests After the estimation of the tracking portfolio weights, the fund manager faces the stability problem of these coefficients. One possible stability test is the "Cumulative Sum Test" defined by Brown et al. [92].

From Equation (4.9), consider the recursive residual terms w_t defined by:

$$w_t = \frac{(y_t - x_t' b)}{\left(1 + x_t' \left(A_t' A_t\right)^{-1} x_t\right)^{\frac{1}{2}}}, \tag{4.17}$$

where x_t' is the observation vectors at time t, A_t the regressor matrix from time 1 to time t, b the vector of regression coefficients, and $y_t - A_t' b$ the value of forecast errors. Its variance is given by:

$$\sigma^2 \left(1 + x_t' \left(A_t' A_t\right)^{-1} x_t\right). \tag{4.18}$$

The CUSUM test is based on the statistics:

$$W_t = \sum_{r=k+1}^{t} \frac{w_r}{s}, \tag{4.19}$$

where w_r is the vector of recursive residual terms, and s is the standard deviation of the regression on the whole sample. If the tracking portfolio weights are constant, then $\mathbb{E}\left(W_t\right) = 0$.

REMARK 4.3 Empirical tests are mitigated. For example, Alexander and Dimitriu [15] compare properties of tracking portfolios based on cointegration and those based on tracking error quadratic minimization. They conclude that there is no significant advantage to use cointegration. However, on other financial data, Dunis and Ho [179] show that the cointegration method reduces the number of rebalancing times. In [417] and [487], such analyses are made on the main French stock index CAC40. The cointegration approach appears to be an efficient method. □

4.2 Benchmark portfolio optimization

As seen in Chapter 3, asset allocation can be based on mean-variance optimization. This allows for the determination of a set of asset class weights. This can induce a "long-term" investment plan, often called the portfolio's *strategic asset allocation.* It can be viewed as *a benchmark* for portfolio optimization. Indeed, the choice of the benchmark is crucial.

However, typically, the portfolio manager has the task of "beating" the benchmark. He may have short-term forecasts which deviate from the long-term forecasts associated to the benchmark. In that case, market timing (active asset allocation) may lead to a divergent portfolio choice. This approach is usually called *tactical asset allocation.* Another example of such divergence from an index or a benchmark is the *core/satellite* approach: according to some specialists, an investor would invest on one hand on an index fund which is considered as the "core" portfolio, and on the other hand choose a skillful manager to add value. This second investment usually called the "satellite" is based on active management.

But, generally, the portfolio manager must choose a portfolio that does not deviate too much from the benchmark. The tracking-error between the portfolio and the benchmark must be controlled. As seen in previous sections, several objective functions can be introduced to measure the risk of the tracking. The most common function is standard deviation. Note that for benchmark portfolio optimization, the tracking error is not necessarily minimized, as it is for index fund management. For instance, it can be fixed at a given "rational" level and other decision criteria can be introduced to determine optimal solutions.

4.2.1 Tracking-error definition

4.2.1.1 Tracking ex-ante and tracking ex-post

The tracking error is defined as the difference between portfolio and benchmark returns. It represents a relative risk of the portfolio with respect to its benchmark. However, as mentioned by Pope and Yadav [411] and Satchell and Hwang [448], the tracking error can be tricky:

- First, the tracking-error can be "anticipated." The risk is measured ex-ante (*tracking ex-ante*). It is a statistical measure. It is a measure of riskiness w.r.t. the benchmark.

- Second, the risk is measured on "realized" risk (*tracking ex-post*). This measures the time series standard deviation of the realized active returns.

For a tracking ex-ante, the fund manager chooses an allocation different from the benchmark's in order to get a better return expectation than the benchmark. To do so, the anticipated function of the tracking, such as standard deviation, may be significantly above 0. For example, for an ex-ante tracking error of 2%, the portfolio returns will fall within $+/-2\%$ of the benchmark with a two-thirds probability. At maturity, ideally, the fund manager will have succeeded in having a realized return at a given level $x\%$ above the benchmark, while the tracking error was almost constant. Unfortunately, the tracking error is often underestimated. Pope and Yadav [411] indicate that autocorrelation of excess returns is a reason for this underestimation. Satchell and Hwang [448] argue that ex-ante and ex-post tracking errors differ when the realized benchmark volatility is high. Haar and van Straalen [279] confirm that ex-post tracking errors are higher than predicted because of changes in volatility, but also because of the concentration in systematic risk factors in a portfolio.

4.2.1.2 Tracking-error, correlation and beta coefficients

We first must distinguish the tracking error from the correlation coefficient and the beta coefficient:

- The correlation coefficient (ρ) measures a "linear risk" between the portfolio return R_P and the benchmark return R_B. For a given standard deviation σ_P of the portfolio return, the correlation coefficient decreases when the specific risk increases. Here, the specific risk indicates the portfolio risk which is independent from the benchmark return.

- The beta coefficient (β_P) measures the risk exposure to the benchmark. For example, if $\beta_P > 1$, then the fund manager incurs a systematic risk w.r.t. the benchmark.

If the objective function on the tracking error T is the variance, denoted by T^2, then we have:

$$\begin{aligned} T^2 &= \sigma^2\left(R_P - R_B\right) = \sigma_B^2 + \sigma_P^2 - 2\rho\sigma_B\sigma_P, \\ &= \sigma_P^2 + \sigma_B^2\left(1 - 2\beta_P\right). \end{aligned} \tag{4.20}$$

Therefore, the tracking-error volatility (TEV) is a function of the total portfolio volatility, of its exposition to the benchmark, and also of the total benchmark volatility. As a consequence, if we focus on excess returns while including benchmark securities, the total risk of the portfolio may be high.

REMARK 4.4 The Sharpe market model allows for the introduction of the relation between the tracking-error, the systematic risk $\beta_P\sigma_B$, and the specific risk $\sigma_{\epsilon P}$.

This model is:

$$R_P = \alpha_P + \beta_P.R_B + \epsilon_P. \tag{4.21}$$

Thus, we have:

$$R_P - R_B = \alpha_P + \left(\beta_P - 1\right).R_B + \epsilon_P, \tag{4.22}$$

where ϵ_P has a standard Gaussian distribution. Then, we deduce:

$$T^2 = \left(\beta_P - 1\right)^2\sigma_B^2 + \sigma_{\epsilon P}^2. \tag{4.23}$$

$\square$

4.2.2 Tracking-error minimization

Several portfolio optimization programs can be introduced to control the tracking-error, depending on what kind of objective function is used. Specific constraints are also often introduced. In what follows, the volatility of the tracking-error is considered, since it is the basic criterion.

Roll [431] has studied two optimizations programs under a constraint on the tracking-error's volatility: minimization of the tracking-error volatility (TEV), and minimization of TEV under a beta constraint.

4.2.2.1 TEV minimization under a mean-return constraint

Consider the same notations as in Chapter 3 for the mean-variance analysis, in particular the definitions of A, B, C, and D (see 3.1.2.1). In addition, denote:

$\mathbf{b}$: the benchmark weight vector

$\mathbf{x}$=$(\mathbf{w} - \mathbf{b})$: the vector of differences between the weights of the portfolio and those of the benchmark

$\overline{R}_B = \mathbf{b}'\overline{R}$: the expected benchmark return

$\sigma_B^2 = \mathbf{b}'V\mathbf{b}$: the variance of the benchmark return

The tracking-error volatility (TEV), denoted also by T, is:

$$T = \sqrt{(\mathbf{w}-\mathbf{b})' \, \mathbf{V} \, (\mathbf{w}-\mathbf{b})} = \sqrt{\mathbf{x}'\mathbf{V}\mathbf{x}}, \tag{4.24}$$

where $\mathbf{V}$ is the variance-covariance matrix of asset returns.

The first optimization problem is the following one:

$$P(1)\colon \begin{cases} \min\limits_{\mathbf{w}} (\mathbf{w}-\mathbf{b})' \, V \, (\mathbf{w}-\mathbf{b}) \\ \text{with } (\mathbf{w}-\mathbf{b})' \, \overline{R} = G, \\ \qquad \mathbf{w}'\mathbf{e} = 1. \end{cases} \tag{4.25}$$

The term G is the excess expected return of the portfolio w.r.t. the benchmark. Roll [431] proves the following result.

PROPOSITION 4.1

1) The optimal solution is given by:

$$\mathbf{w} = \mathbf{b} + D\left(\mathbf{q}_1 - \mathbf{q}_0\right), \tag{4.26}$$

where

$$D = \frac{G}{\overline{R}_1 - \overline{R}_0} \text{ and } \mathbf{q}_0 = \mathbf{V}^{-1}\frac{\mathbf{B}}{C}, \; \mathbf{q}_1 = \mathbf{V}^{-1}\frac{\mathbf{R}}{A}. \tag{4.27}$$

The parameter coefficient D can be viewed as a relative performance coefficient. The portfolio $\mathbf{q}_0$ is the mvp, and $\mathbf{q}_1$ is also a mean-variance efficient portfolio with:

$$\overline{R}_0 = \frac{A}{C}, \; \overline{R}_1 = \frac{B}{A},$$
$$\sigma_0^2 = \frac{1}{C}, \; \sigma_1^2 = \frac{B}{A^2}.$$

2) The tracking-error variance and the total variance of optimal solutions from Problem $P(1)$ are given by:

$$T^2 = D^2\left(\sigma_1^2 - \sigma_0^2\right),$$
$$\sigma^2 = \sigma_B^2 + T^2 + 2D\sigma_0^2\left(\overline{R}_B/\overline{R}_0 - 1\right). \tag{4.28}$$

REMARK 4.5 Optimal portfolios given in relation (4.26) are the sums of the benchmark and of a deviation term. This one depends only on two special mean-variance efficient portfolios and on the anticipated excess return, which is independent from the benchmark. The tracking-error volatility (relation (4.28)) is an increasing linear function w.r.t. the anticipated excess return, which does not depend on the benchmark.

□

DEFINITION 4.3 *The set of optimal solutions associated to Problem $P(1)$ is generated by the variation of the excess return G. In risk-return space, this set is called the relative frontier.*

REMARK 4.6 Most of the time, the benchmark is not mean-variance efficient, as empirically shown by Grinold [272]. If the benchmark is efficient, then all optimal solutions of Problem $P(1)$ are also mean-variance efficient. Otherwise, they are not. □

Portfolio solutions of Problem $P(1)$ have two shortcomings:

- First, generally they do not mean-variance dominate the benchmark.
- Second, their beta coefficients w.r.t. the benchmark are higher than 1.

These properties are examined in what follows.

A) Mean-variance dominance.

- Case 1: $\overline{R}_B > \overline{R}_0$. Solutions of Problem $P(1)$ *do not dominate the benchmark.* This case is illustrated by the following figure.

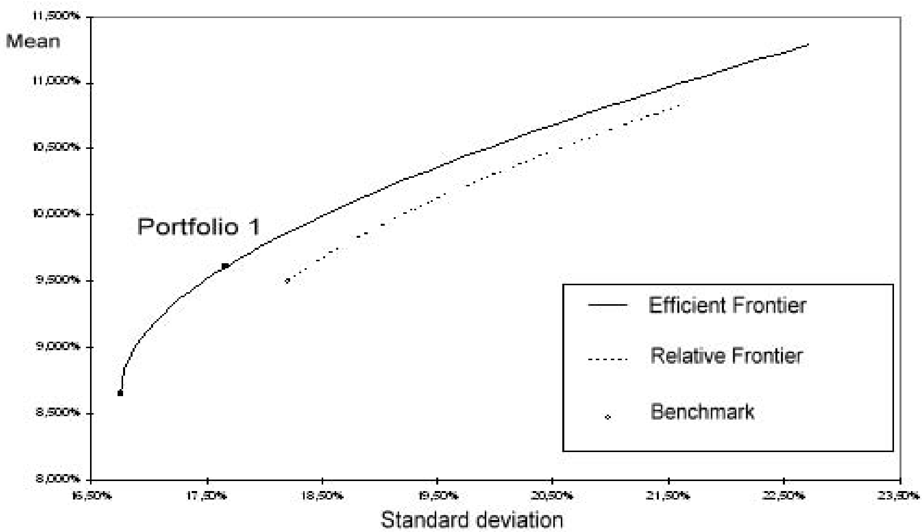


FIGURE 4.4: Efficient and relative frontiers, $R_B > R_0$

In that case, the benchmark has an expected return above the mvp expected return. Then, if the excess return of a portfolio w.r.t the benchmark is non-negative ($G > 0$), the portfolio total risk is higher than the benchmark total

risk (see relation (4.28)). If a riskless asset is introduced, such property is still satisfied for benchmark having an expected return higher than the riskless one. Therefore, for this first case which is the most usual, none of the solutions of Problem $P(1)$ mean-variance dominate the benchmark (which also does not dominate them).

- Case 2 : $\overline{R}_B < \overline{R}_0$. Some of the optimal solutions dominate the benchmark. For small tracking error volatility, the total risk is smaller than the benchmark total risk. However, for a higher excess return G, optimal solutions no longer dominate the benchmark. These properties are illustrated by the following figure, which includes both mean-variance and relative efficient frontiers.

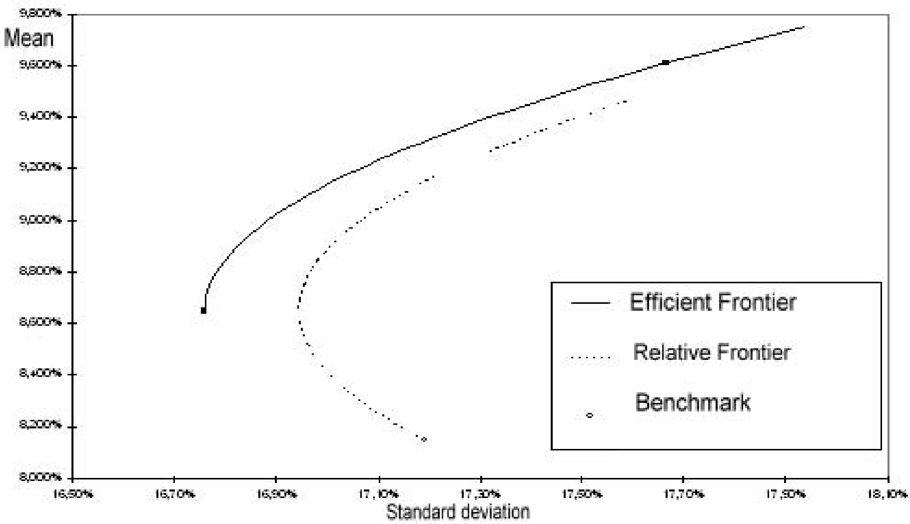


FIGURE 4.5: Efficient and relative frontiers, $R_B < R_0$

Indeed, the term D measures a relative gain when taking a sufficient tracking-error in order to generate an excess return. When D is small (the portfolio is close to the benchmark), a first-order Taylor's expansion leads to the following approximation:

$$\sigma_P^2 \simeq \sigma_B^2 + 2D\left(\overline{R}_B/\overline{R}_0 - 1\right). \tag{4.29}$$

The portfolio total risk σ_P^2 is smaller than the benchmark total risk σ_B^2.

This explains why the portfolio manager gets a portfolio which dominates the benchmark (according to mean-variance criterion).

B) Beta coefficients w.r.t. the benchmark.

The beta coefficients of optimal solutions w.r.t. the benchmark are given by:

$$\beta_P = 1 + D\left(\frac{\sigma_0^2}{\sigma_B^2}\right)\left(\overline{R_B}/\overline{R_0} - 1\right). \tag{4.30}$$

Consider the usual case when the benchmark has an expected return higher than the expected return of the mvp ($\overline{R_B} > \overline{R_0}$). Then, $\beta_P > 1$. Therefore, the systematic risk w.r.t. the benchmark is significant, as soon as the portfolio manager has a *benchmark-timing strategy.*

4.2.2.2 TEV minimization under mean-return and beta constraints

In order to mitigate the previous drawbacks, Roll [431] proposes a relative optimization problem with an additional constraint on the portfolio beta. For a given portfolio exposition to the benchmark (the beta), we can search for portfolios which minimize the tracking error. The interesting point is to get optimal solutions with beta smaller than 1, according to the given constraint.

The new optimization problem is:

$$P(2){:}\begin{cases} \min\limits_{\mathbf{w}} (\mathbf{w}-\mathbf{b})' V (\mathbf{w}-\mathbf{b}) \\ \text{with } (\mathbf{w}-\mathbf{b})' \overline{R} = G, \\ \mathbf{w}'\mathbf{e} = 1, \\ \dfrac{\mathbf{w}'\mathbf{V}\mathbf{b}}{\sigma_B^2} = \beta. \end{cases} \tag{4.31}$$

PROPOSITION 4.2

The solution of Problem $P(2)$ is given by:

$$\mathbf{w}^* = \mathbf{b} + \nu\mathbf{b} + \mathbf{V}^{-1}\left(\gamma\overline{\mathbf{R}} + \delta\mathbf{e}\right), \tag{4.32}$$

where the Lagrange multipliers are given by:

$$\gamma = \frac{G\left(1 - C\sigma_B^2\right) - \sigma_B^2(\beta - 1)(B - CR_B)}{(A - BR_B)(1 - C\sigma_B^2) - (B - CR_B)(R_B - B\sigma_B^2)}, \tag{4.33}$$

$$\delta = \frac{\sigma_B^2(\beta - 1)(A - BR_B) - G\left(R_B - B\sigma_B^2\right)}{(A - BR_B)(1 - C\sigma_B^2) - (B - CR_B)(R_B - B\sigma_B^2)}, \tag{4.34}$$

$$\nu = -\gamma B - \delta C. \tag{4.35}$$

Roll [431] proves that the optimal solution is a convex combination of the three portfolios $\mathbf{q}_0$, $\mathbf{q}_1$, and the benchmark $\mathbf{b}$.

The following figure presents three "beta frontiers."

Note that for some fixed beta values (for example $\beta < 1$), some solutions of Problem $P(2)$ mean-variance dominate the relative frontier associated to

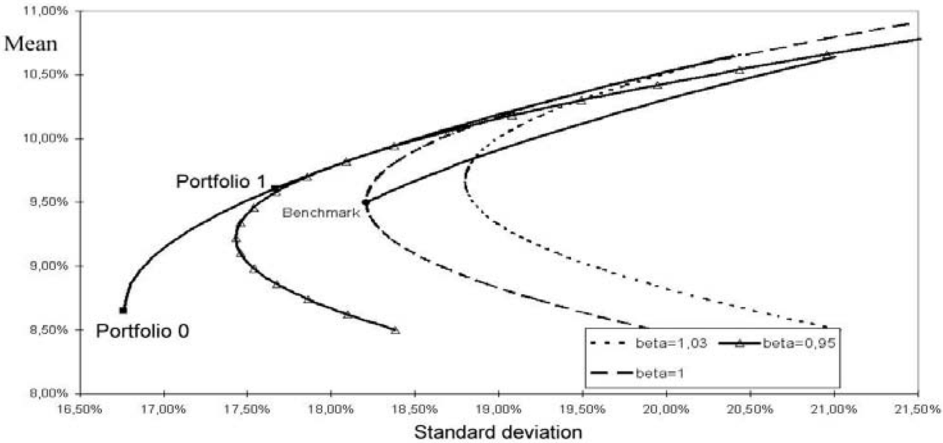


FIGURE 4.6: Efficient, relative, and beta frontiers

Problem $P(1)$. Nevertheless, these portfolios have tracking-error volatilities higher than those of the relative frontier.

REMARK 4.7 By fixing beta values smaller than 1, the benchmark-timing risk is limited. However, for $\beta \neq 1$, the beta frontier does not contain the benchmark, which may be a drawback.

$\square$

4.2.2.3 Mean-variance optimization under TEV constraint

As mentioned by Roll [431], excess return optimization leads to optimal portfolios with systematically higher total risk than the benchmark. Thus, we can search for a criterion which controls both the total risk and the relative risk (TEV).

For example, Bertrand *et al.* [61] consider that investors (or fund managers) maximize a mean-variance criterion under a tracking-error volatility constraint. In this framework, the manager has to solve the following optimization problem:

$$P(3): \begin{cases} \max \ \mathbf{w}'\mathbf{R} - \frac{\phi}{2}\mathbf{w}'\mathbf{V}\mathbf{w} \\ \text{with } (\alpha - b)' \, \mathbf{V} \, (\alpha - b) = T^2, \\ \mathbf{w}'\mathbf{e} = 1. \end{cases} \tag{4.36}$$

The parameter ϕ is the marginal substitution rate between the return and the variance. It can be viewed as an aversion w.r.t. the variance. In what follows, we call it the “variance aversion.”

PROPOSITION 4.3
The optimal solution is given by (see Bertand et al. [61]):

$$\begin{cases} \mathbf{w}^{**} = \mathbf{b} + \frac{1}{\phi+\lambda}\left(-\phi\mathbf{b} + \mathbf{V}^{-1}\left(\overline{\mathbf{R}} - \mu\mathbf{e}\right)\right) \\ \lambda = -\phi + \frac{\sqrt{\phi^2\left(\sigma_B^2-\sigma_0^2\right)+2\phi(R_0-R_B)+\frac{d}{C}}}{T}, \\ \mu = \frac{B-\phi}{C}. \end{cases}$$

REMARK 4.8 The set of portfolio solutions of the previous Problem $P(3)$ is the same as the set of portfolio solutions of Roll program, as soon as the expected rate of return of the benchmark is greater than that of the minimum variance portfolio.

□

For a fixed variance aversion ϕ, the ϕ-frontier associated to Problem $P(3)$ is generated when the TEV is varying.

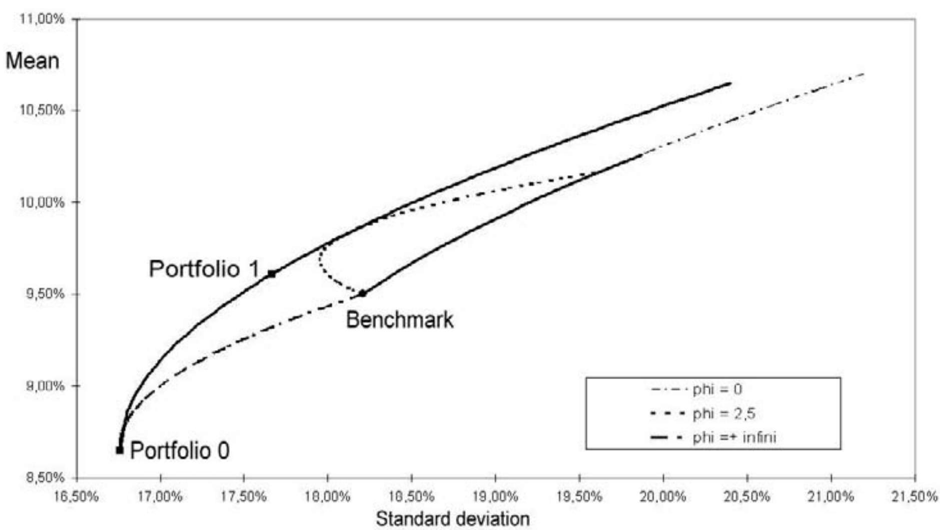


FIGURE 4.7: Efficient, relative, and Φ frontiers

When the variance aversion ϕ increases, the ϕ-frontiers move to the left (area of portfolios with small total risk). If the variance aversion is high, the portfolio manager chooses portfolios with a total risk smaller than the benchmark total risk. If the variance aversion goes to infinity, then the optimal portfolio converges to the mvp, if the TEV is sufficiently high. For some values of ϕ (for example $\phi = 2.5$), some optimal portfolios mean-variance dominate the benchmark. Note also that all these frontiers contain the benchmark.

REMARK 4.9 When $\phi = 0$, the set of optimal solutions $P(3)$ is equal to the set of solutions of Problem $P(1)$. The relative frontier corresponds to investors who are "neutral" w.r.t. the total risk ($\phi = 0$). Nevertheless, we have to choose an aversion risk level. For example, we can consider values of ϕ such that there exists a tangency point between the ϕ-frontier (solution of $P(3)$) and the mean-variance frontier. An optimal portfolio can be chosen on that curve, according to the TEV. The set of optimal solutions of Problems $P(2)$ and $P(3)$, which have mean returns higher than the mvp return, are equal: any solution of $P(2)$, associated to a pair (G, β), is equal to one and only one solution of $P(3)$, associated to a pair (ϕ, T).

▯

Example 4.1

Consider a fund manager who chooses a TEV equal to 3.00%. For the parameter values considered in this example, we have:

1) For $\phi = 2.4145$, the portfolio solution of Problem $P(3)$ has the following characteristics:

Standard deviation	Mean	TEV	Beta
18.00%	9.70%	3.00%	0.975

This portfolio mean-variance dominates the benchmark. Its beta is smaller than 1. Therefore, this portfolio is closer to the benchmark than the corresponding portfolio belonging to the relative frontier.

2) Consider now the portfolio belonging to the relative frontier with the same TEV. Its characteristics are:

Standard deviation	Mean	TEV	Beta
19.24%	10.02%	3.00%	1.045

The fund manager may choose to lose 0.32% in the mean return in order to reduce the total volatility (−1.24%) while also reducing the exposure beta (from $1.045 > 1$ to $0.975 < 1$).

3) We can also consider the TEV-frontier with TEV=3.00% (the constant-TEV-frontiers are ellipses in the risk/return space). Then, we can select the portfolio belonging to this curve with the same total risk as the benchmark. Its characteristics are:

Standard deviation	Mean	TEV	Beta
18.21%	9.787%	3.00%	0.986

Contrary to the solution of Problem $P(3)$ with $\phi = 2.4145$, this porfolio does not mean-variance dominate the benchmark.

The following figure shows the comparison between the three criteria and the associated solutions.

▯

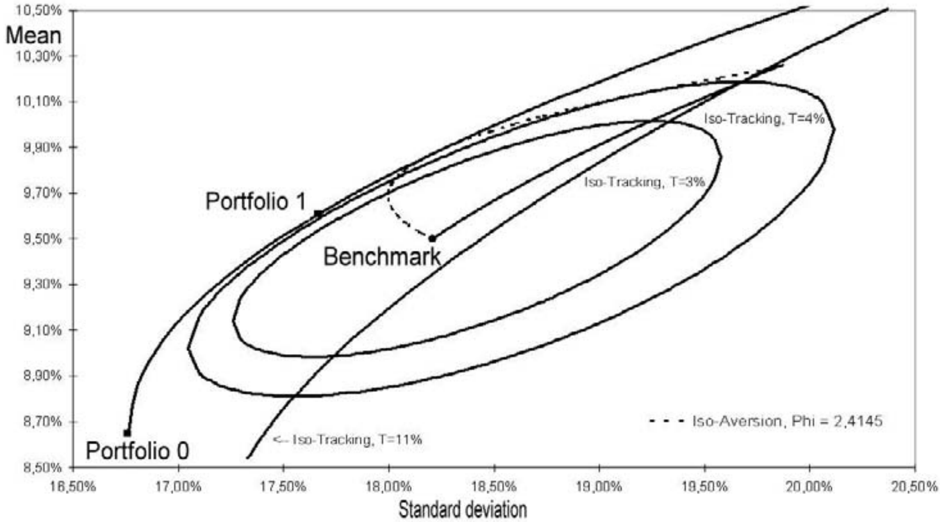


FIGURE 4.8: Frontiers and iso-tracking curves

4.3 Further reading

As examined in Sorenson et al. [474], investors and fund managers have to allocate between active and passive management.

Strategic allocation consists of defining the orientation of the investment:

By determining the risk aversion and horizon in order to choose the benchmark;

By choosing the portfolio composition: asset class allocation among cash, fixed-income products (domestic or international) and stocks (by sectors or styles), alternative investments, *etc*; and,

By fixing the tolerance of the fund manager with respect to the benchmark.

Tactical allocation is more concerned with the effective management process. It consists of searching for the best way to achieve the previous goal:

By choosing the weighting asset class with respect to the benchmark. This can be done by market timing (which induces the determination of the exposure to the benchmark), and by asset allocation among sectors, lands, *etc.*

By using asset valuation models.

Main Index tracking methods are based on statistical replication which goes back to the 70s. Many articles have examined this method. Rudd [441] examines the selection of passive portfolios. Rudd and Rosenberg [442] consider portfolio optimization problems and their realistic implementation. Bamberg and Wagner [40] search for robust regression estimators to replicate equity indexes. Dash et al.[148] attempt to measure the efficiency and costs of such portfolio optimization.

Evolutionary algorithms are appropriate candidates for being able to continuously track the movement of the optimum through time. In order to solve large scale optimization problems with constraints, "metaheuristic" methods have been introduced, *e.g.*, by Derigs and Nickel [163] for index tracking. They are based on neighborhood methods and evolutionary algorithms such as genetic algorithms. These latter methods are also used in other applications in finance, as shown by Bauer [51]. Conditions insuring the convergence of the

threshold-accepting algorithm are also provided in Althöfer and Koschnick [21]. Applications of this algorithm in econometrics is provided in Winker [505].

An overview of some different benchmark portfolio optimization procedures is given in Wagner [500] (see also Wagner [499]). Other references are: Franks [241] and Clarke *et al.* [124] who study tracking errors and tactical asset allocation, and, Jorion [310] who proves that the set of tracking-error volatility constrained portfolios is an ellipse on the mean-variance plane.

Gaivoronski and Krylov [250] present several portfolio selection algorithms in a dynamic setting which take into account different risk/target measures. They search for the best tradeoff between tracking precision and rebalancing frequency in order to control costs.

Chapter 5

Portfolio performance

Competition among financial institutions has led to the performance analysis which examines the qualities of managers who use active investment strategies. The investment process is overall evaluated.

The fundamental question is:

Does a given manager provide an actual value-added service with respect to a simple index replication or to a given benchmark?

As a by-product, the performance analysis is a test of the market efficiency which is based on different kinds of information, as introduced by Fama [217].

An active manager tries to take opportunities from particular information (public or private) which may not be reflected by market prices.

To analyze a manager's performance, it is necessary:

- First, to introduce specific performance measures, taking account of different types of risk. Therefore, several risk measures can be considered in order to evaluate the risk associated to the performance.

- Second, to attribuate this performance:

  - What are the exact contributions of each investment decision to the overall portfolio performance?
  - Is the manager skillful or lucky?
  - Compared with the benchmark, does the performance come from stock picking or from market timing?

- Finally, to examine the performance consistency? More precisely, is past performance a good indicator of future performance?

5.1 Standard performance measures

In order to ensure that fund comparisons are legitimate:

- First, funds must have the same nature.
- Second, a standardization of performance measures and their results must be introduced.

The AIMR (*Association for Investment Management and Research*) provided such standards, further developed by the GIPS (*Global Investment Performance Standard*). They focus mainly on equities and fixed-income securities. Generally, two measures are considered: a measure of total risk, such as the standard deviation, and a measure of market risk, such as the beta. The first one is "absolute," the second one is "relative."

For the latter measure, we have to evaluate the sensitivity of securities to the market. This can be done by using the fundamental CAPM introduced by Sharpe ([462],[463]).

5.1.1 The Capital Asset Pricing Model

Suppose that there exists a riskless asset with return R_f. As seen in Chapter 3, under the Markowitz assumptions, the two-fund separation theorem is valid: any efficient portfolio P is a combination of the riskless asset and of the market portfolio M, which corresponds to the point of tangency between the two efficient frontiers (with and without the riskless asset).

Then, we have:

$$\overline{R}_P = xR_f + (1-x)\overline{R}_M \text{ and } \overline{R}_P - R_f = (1-x)\left(\overline{R}_M - R_f\right). \tag{5.1}$$

The choice of x depends on the risk aversion. Therefore, its variance is equal to:

$$\sigma_P = (1-x)\sigma_M. \tag{5.2}$$

Consequently,

$$\overline{R}_P = R_f + \sigma_P \left(\frac{\overline{R}_M - R_f}{\sigma_M}\right). \tag{5.3}$$

In the Markowitz plane $(\sigma(R), \overline{R})$, this equation defines the efficient frontier, which is also called the *capital market line.*

Following Sharpe's demonstration, at equilibrium the prices of assets are such that the market portfolio is made up of all assets in proportion to their market capitalizations (in practice, a stock exchange index). Then, consider

the portfolio P_i (not necessarily efficient) with a proportion x invested on asset i, and $(1-x)$ invested on M. We have:

$$\overline{R}_P = x\overline{R}_i + (1-x)\overline{R}_M, \tag{5.4}$$

$$\sigma_P = \left[x^2 + \sigma_i^2 + (1-x)^2\sigma_M^2 + 2x(1-x)\sigma_{iM}\right]^{1/2}. \tag{5.5}$$

Consider now the curve of all possible portfolios P_i when x is varying. The coefficient of the tangent to this curve at the point associated to x is given by:

$$\frac{\partial \overline{R}_P}{\partial \sigma_P} = \frac{\partial \overline{R}_P/\partial x}{\partial \sigma_P/\partial x} = \frac{\left(\overline{R}_i - \overline{R}_M\right)\sigma_P}{x\left(\sigma_i^2 + \sigma_M^2 - 2\sigma_{iM}\right) + \sigma_{iM} - \sigma_M^2}. \tag{5.6}$$

Since for the market portfolio M we have $x = 0$, we deduce from Equation (5.3) that:

$$\frac{\left(\overline{R}_i - \overline{R}_M\right)\sigma_M}{\sigma_{iM} - \sigma_M^2} = \left(\frac{\overline{R}_M - R_f}{\sigma_M}\right),$$

which is equivalent to:

$$\overline{R}_i - R_f = \beta_i(\overline{R}_M - R_f) \text{ with } \beta_i = \sigma_{iM}/\sigma_M^2. \tag{5.7}$$

The coefficient beta represents the systematic risk which is due to exposition to the market variations. In the plane $(\beta_M(R), \overline{R}])$, the previous equation defines a straight line, the so-called *security market line*. At equilibrium, all assets (thus all portfolios) are located on this line.

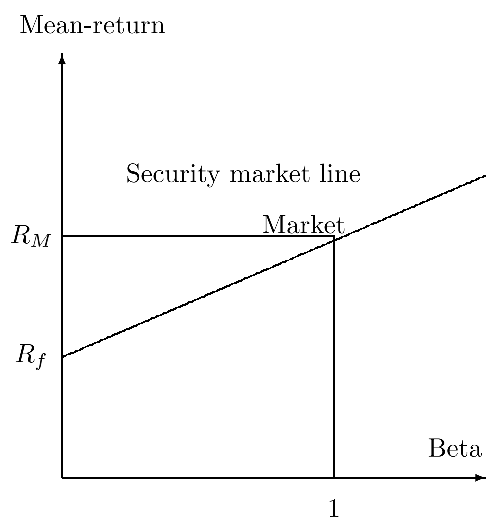


FIGURE 5.1: Security market line

Note also that, at equilibrium, the market portfolio is optimal, which is in favor of passive management based on index funds. As shown by Black

[73], the CAPM is still valid without the riskless asset, which is replaced by a zero-beta portfolio Z:

$$\overline{R}_i - \overline{R}_Z = \beta_i(\overline{R}_M - \overline{R}_Z) \text{ with } \beta_i = \sigma_{iM}/\sigma_M^2.$$

Taxes can also be taken into account, as in Brennan [87]:

$$\overline{R}_i - R_f - T(D_i - R_f) = \beta_i(\overline{R}_M - R_f - T(D_M - R_f)) \text{ with } T = \frac{T_d - T_g}{1 - T_g},$$

where T_d denotes the average taxation rate for dividends, T_g denotes the average taxation rate for capital gains, D_i and D_M denote respectively the dividend yields of asset i and market portfolio M.

The CAPM highlights the relationship between the excess mean return (the "reward") and the exposure coefficient beta (the "risk"). Other factors can also be introduced to estimate the excess return, as shown in the next section on performance decomposition.

5.1.2 The three standard performance measures

These performance measures are based on the previous properties: Relation (5.3) which defines the capital market line and relation (5.7) which defines the security market line. The higher these performance measures, the more interesting the portfolio.

5.1.2.1 The Sharpe measure

This measure is based on the the capital market line. For all efficient portfolios, we have the following equality between (excess reward)/(total risk) ratios:

$$\frac{\overline{R}_P - R_f}{\sigma(R_P)} = \frac{\overline{R}_M - R_f}{\sigma(R_M)}. \tag{5.8}$$

The previous ratio is the slope of the capital market line. As for Markowitz portfolio optimization, we search for such portfolios, since this is equivalent to the maximization of the ratio $RS = \frac{\overline{R}_P - R_f}{\sigma(R_P)}$. Thus, as proposed by Sharpe [464], this can be actually considered as a performance measure. It is equal to the excess mean return (w.r.t. the riskless asset) and the measure ot total risk (the standard deviation), and is defined as the "*reward-to-variability ratio.*" For example, the manager can check if the excess mean return of the portfolio is sufficient to compensate a higher risk than the market portfolio. If the portfolio is well-diversified, its Sharpe ratio is close to the market portfolio's (see also comments in Sharpe [466]).

5.1.2.2 The Treynor measure

The Treynor's ratio [494] is directly based on the CAPM:

$$RT = \frac{\overline{R}_P - R_f}{\beta_P}. \tag{5.9}$$

It can also be viewed as a reward-to-risk ratio where the "risk" is the exposition to market risk. At equilibrium, this ratio is constant and equal to $\overline{R}_M - R_f$. The Treynor ratio allows us to evaluate the performance of a well-diversified portfolio, since it only involves the systematic risk. It can be used to examine performance of portfolio which is only a part of the investor's assets. Due to previous diversification, the investor takes care only of the systematic risk.

5.1.2.3 The Jensen measure

This measure is also based on the CAPM, but mainly when it is not satisfied. Indeed, consider the difference α_P between the mean excess return of portfolio P and the return explained by the CAPM. Then:

$$\alpha_P = \left(\overline{R}_P - R_f\right) - \beta_P \left(\overline{R}_M - R_f\right). \tag{5.10}$$

Consequently:

$$\overline{R}_P - R_f = \alpha_P + \beta_P \left(\overline{R}_M - R_f\right). \tag{5.11}$$

The coefficient α_P is called the performance measure introduced by Jensen [300]. Therefore, the manager searches for portfolios with $\alpha_P > 0$. For ex-post measures, the manager will be judged skillful if the coefficient α_P is significantly above 0.

If $\alpha_P = 0$, the return of portfolio P is at equilibrium and the manager's forecasts have not beat the market performance: the portfolio has the same alpha as any combination of the riskless asset and the market portfolio. Unlike the Sharpe and Treynor ratios, the Jensen measure contains the benchmark itself. As with the Treynor ratio, it only takes account of the systematic risk. Due to the particular form of alpha, only portfolios with the same risk beta can be compared. Otherwise, we can consider, for example, the Black-Treynor ratio:

$$RBT = \frac{\alpha_P}{\beta_P}.$$

REMARK 5.1 Previous ratios are defined *ex-ante*. However, ex-post formulas can be also introduced, as shown by Jensen [300]. Consider the market model:

$$R_{Pt} = \gamma_P + \beta_P R_{Mt} + \varepsilon_{Pt}, \tag{5.12}$$

where ε_{Pt} has zero expectation, is independent from R_{Mt}, and:

$$Cov(\varepsilon_{Pt}, \varepsilon_{Ps}) = 0, \forall t \neq s.$$

The return expectation of portfolio P is given by:

$$\overline{R}_P = \gamma_P + \beta_P \overline{R}_M.$$

Thus:

$$R_{Pt} = \overline{R}_P + \beta_P \left(R_{Mt} - \overline{R}_M\right) + \varepsilon_{Pt}.$$

Finally:

$$R_{Pt} - R_f = \alpha_P + \beta_P \left(R_{Mt} - R_f\right) + \varepsilon_{Pt}. \tag{5.13}$$

Assume that R_f and β_P are constant. Then, we can get an unbiased estimation of the coefficient α_P by using a least squares regression. The estimators of the unknown parameters α_P and β_P are given by

$$\widehat{\beta}_P = \frac{\widehat{Cov\left(R_M - R_f, R_P - R_f\right)}}{\sigma^2\left(\widehat{R_M - R_f}\right)},$$
$$\widehat{\alpha}_P = \left(\widehat{R}_P - R_f\right) - \widehat{\beta}_P\left(\widehat{R}_M - R_f\right),$$

where the symbol $\widehat{X}$ denotes the empirical mean of the sample which contains T observations of the random variable α. The significance of the estimator $\hat{\alpha}_P$ can be based on a Student test on the hypothesis $H_0 : \hat{\alpha}_P = 0$. The *ex-post* Sharpe and Treynor ratios are given by:

$$\widehat{RS} = \frac{\widehat{R}_P - R_f}{\widehat{\sigma}\left(R_P\right)} \text{ and } \widehat{RT} = \frac{\widehat{R}_P - R_f}{\widehat{\beta}_P}. \tag{5.14}$$

□

REMARK 5.2 Jensen [300] considers the following beta specification:

$$\beta_{Pt} = \mathbb{E}\left[\beta_P\right] + \varepsilon_{Pt},$$

where ε_{Pt} is a white noise. The manager has a mean beta target $\mathbb{E}\left[\beta_P\right]$. In this framework, a particular estimation model of the alpha coefficient α_P is introduced (see [300]). The manager can adjust the beta from the global market forecasts.

□

5.1.2.4 Performance measures comparison

From Equation 5.11, relations between all these measures can be deduced.

PROPOSITION 5.1
(Relations between Sharpe, Treynor, and Jensen measures)

- The Sharpe ratio is also equal to:

$$RS = \frac{\overline{R}_P - R_f}{\sigma(R_P)} = \frac{\alpha_P}{\sigma(R_P)} + \frac{\rho(R_p, R_M)}{\sigma(R_M)}\left(\overline{R}_M - R_f\right). \tag{5.15}$$

For any efficient portfolio P, we have: $\rho(R_p,\ R_M) = 1$*. Therefore, we get:*

$$RS = \frac{\alpha_P}{\sigma(R_P)} + \frac{(\mathbb{E}[R_M] - R_f)}{\sigma(R_M)}. \tag{5.16}$$

- The Treynor ratio is simply equal to:

$$RT = \frac{\mathbb{E}[R_P] - R_f}{\beta_P} = \frac{\alpha_P}{\beta_P} + (\mathbb{E}[R_M] - R_f). \tag{5.17}$$

- If the portfolio is well-diversified, we have:

$$\rho(R_p,\ R_M) \simeq 1. \text{ Thus: } \beta_P \simeq \frac{\sigma_P}{\sigma_M}. \text{ Then: } RS \simeq \frac{RT}{\sigma_M}.$$

The following example illustrates properties of the three performance measures.

Example 5.1
Suppose that the fund manager has private information about a given firm. The portfolio denoted by A contains only this stock. We have:

$$R_A = \alpha_A + R_f + \beta_A(R_M - R_f) + \varepsilon_A.$$

We assume that this information is sufficiently relevant such that:

$$\underset{(4.9\ \%)}{\mathbb{E}[R_A]} = \underset{(0.5\ \%)}{\alpha_A} + \underset{(2\ \%)}{R_f} + \underset{(0.687)}{\beta_A}\left(\underset{(5.49\ \%)}{\mathbb{E}[R_M]} - R_f\right).$$

The total risk is assumed to be equal to 6% and the specific risk equal to 4.72%. These risks are assumed to not be modified by the information.

A second fund manager has a well-diversified portfolio B due to the "good" knowlegde of many stocks with total risk equal to 8.71% and the specific risk equal to 4.73%. We have:

$$R_B = \alpha_B + R_f + \beta_B(R_M - R_f) + \varepsilon_B.$$

with:

$$\underset{(7.24\ \%)}{\mathbb{E}\left[R_B\right]} = \underset{(0.5\ \%)}{\alpha_B} + \underset{(2\ \%)}{R_f} + \underset{(1.359)}{\beta_B}\left(\underset{(5.49\ \%)}{\mathbb{E}\left[R_M\right] - R_f}\right). \tag{5.18}$$

The next figure illustrates these assumptions. Points A' and B' correspond to portfolios A and B at equilibrium (*i.e.*, with 0 alpha).

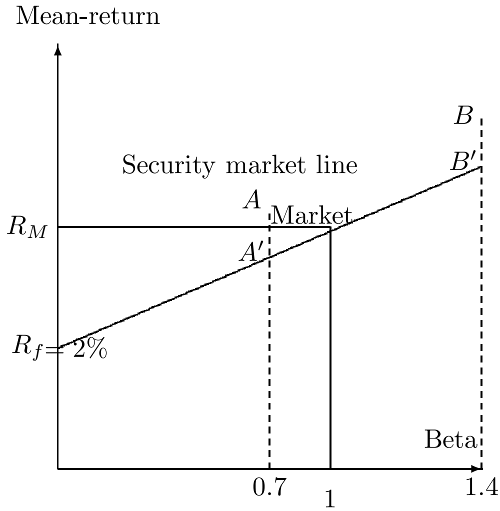


FIGURE 5.2: SML and portfolios A, B, A' and B'

The following figure represents the efficient frontier without riskless asset and the CML.

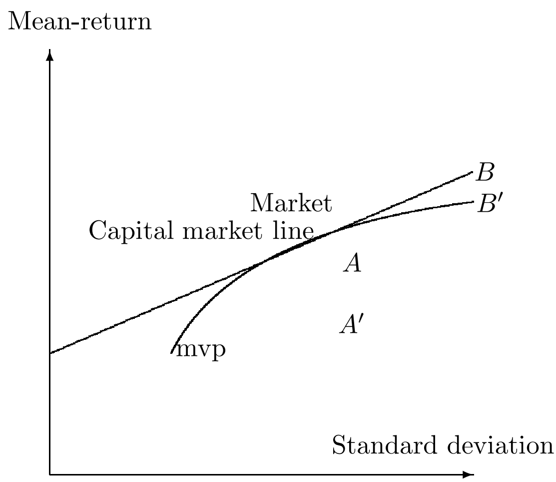


FIGURE 5.3: Capital market line

The well-diversified fund B' is assumed to be efficient, while the fund A', which contains only one stock, is not on this frontier. Using the alpha measure, funds A and B cannot be distinguished since both fund managers have the same excess return w.r.t. the CAPM (equal to 0.5%). For this special case, the Jensen measure does not allow for the comparison of these two portfolios, having too different risks.

As mentioned by Modigliani and Pogue [392], the Treynor ratio allows us to reduce the bias due to the leverage effect which is contained in the Jensen alpha. Indeed, fund A has an excess return per systematic risk unit which is higher than fund B. Assuming that we can borrow the riskless asset, the leverage effect allows us to reach A^* (from A), which has the same beta as fund B but a higher mean return. The following figure shows A, A^*, and B. Funds A and A^* have same Treynor ratio (equal to 4.22%) which is higher than fund B (3.86%).

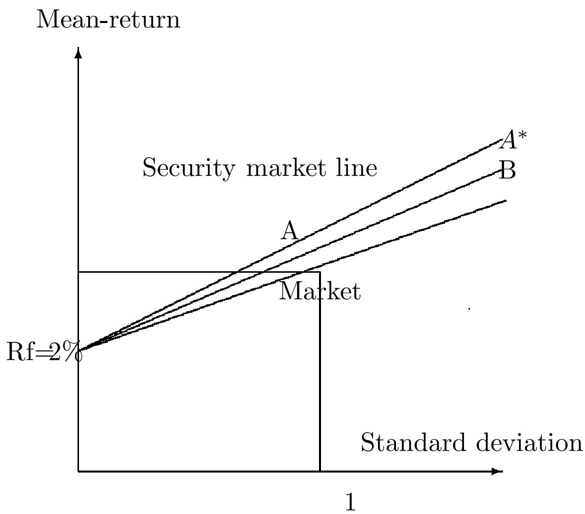


FIGURE 5.4: Capital market line and leverage effect

The Sharpe ratio applied to funds A and B allows us to take account of the diversification. The Sharpe ratio of fund A is equal to 0.4833, while for fund B it is equal to 0.6025. Therefore, the manager who chooses the more diversified portfolio has a higher performance ranking.

□

PROPOSITION 5.2

(A condition ensuring the same ranking by Sharpe and Treynor ratios)

Consider two funds A and B. Their rankings w.r.t. Sharpe and Treynor ratios are equivalent if they have the same correlation coefficient with the

market portfolio:

$$(RT_A > RT_B \iff RS_A > RS_B) \text{ if } \rho_{AM} = \rho_{BM}. \tag{5.19}$$

PROOF We have:

$$RT_A > RT_B \iff \frac{R_A - R_f}{\beta_A} > \frac{R_B - R_f}{\beta_B},$$
$$\iff \frac{\sigma_M}{\rho_{AM}} \frac{R_A - R_f}{\sigma_A} > \frac{\sigma_M}{\rho_{BM}} \frac{R_B - R_f}{\sigma_B} \iff \frac{RS_A}{\rho_{AM}} > \frac{RS_B}{\rho_{BM}}.$$

□

REMARK 5.3 If $RT_A > RT_B$ and $\rho_{AM} > \rho_{BM}$, then $RS_A > RS_B$. Otherwise, no general comparison is possible. Using the market model, we have:

$$\sigma_A^2 = \beta_A^2 \sigma_M^2 + \sigma_{\varepsilon_A}^2 \iff \sigma_A^2 = \rho_{AM}^2 \sigma_A^2 + \sigma_{\varepsilon_A}^2 \iff \rho_{AM}^2 = 1 - \frac{\sigma_{\varepsilon_A}^2}{\sigma_A^2}. \tag{5.20}$$

Thus, the higher the correlation coefficient, the smaller the specific risk for a given total risk. Then, if fund A has a higher Treynor ratio than fund B and a smaller (specific risk)/(total risk) ratio, then fund A has a higher Sharpe ratio than fund B.

□

PROPOSITION 5.3

(A condition ensuring the same ranking by Sharpe and Jensen measures)

If $\sigma_A < \sigma_B$ and $\rho_{AM} > \rho_{BM}$, then:

$$\alpha_A > \alpha_B \iff RS_A > RS_B. \tag{5.21}$$

PROOF The Sharpe ratios are given by:

$$RS_A = \frac{\alpha_A}{\sigma_A} + \frac{\beta_A}{\sigma_A} \left(\mathbb{E}\left[R_M\right] - R_f \right), \; RS_B = \frac{\alpha_B}{\sigma_B} + \frac{\beta_B}{\sigma_B} \left(\mathbb{E}\left[R_M\right] - R_f \right).$$

Therefore:

$$RS_A > RS_B \iff \alpha_A - \frac{\sigma_A}{\sigma_B}.\alpha_B > RS_M \left(\rho_{BM} - \rho_{AM} \right) \sigma_A.$$

Using the assumption $\sigma_A < \sigma_B$ and $\rho_{AM} > \rho_{BM}$, we deduce:
$\alpha_A > \alpha_B \iff RS_A > RS_B$.

□

REMARK 5.4 The Sharpe ratio does not modify the Treynor ratio ranking $RT_A > RT_B$ if $\rho_{AM} > \rho_{BM}$. In order to get such general relation between the Sharpe ratio and the Jensen measure, we have to introduce an additional condition on the total risk: $\sigma_A < \sigma_B$.

□

5.1.2.5 Roll's criticism

According to Roll [430]), it is very difficult to determine the true market portfolio. Moreover, the market portfolio has to be mean-variance efficient. But, the true market portfolio cannot be observed since it must contain all the risky assets, even those which are not traded. Since a stock exchange index has to be substituted, the empirical studies depend on this choice. The CAPM is validated if this index is mean-variance efficient.

Therefore, Roll's criticism concerns performance measures based on the CAPM, such as the Treynor ratio and the Jensen measure. Using a portfolio which is not the true market portfolio may lead to estimation errors in the betas. However, the measurement errors can be corrected, as proposed by Shanken [461]. The Sharpe ratio avoids this problem, since it is based on the total risk instead of the beta. However, if we compare the Sharpe ratio of a given portfolio to the Sharpe ratio of an index, this comparison obviously depend on the choice of index. In addition, some statistical problems may also appear when estimating the Sharpe ratio.

REMARK 5.5 What is the precision of Sharpe ratio estimation when using financial data? Lo [359] analyzes the statistical distribution of the Sharpe ratio, assuming different properties on the financial process which generates the data. If portfolio returns are supposed to be independent and identically distributed (*iid*) then, from central limit theorem, the asymptotic distribution of the Sharpe ratio is such that:

$$\sqrt{T}\left(\widehat{RS} - RS\right) \overset{a}{\sim} \mathcal{N}\left(0, 1 + \frac{1}{2}RS^2\right).$$

Table 5.1 indicates some standard deviation values of the asymptotic Sharpe ratio $\left(SE\left(\widehat{RS}\right) \overset{a}{=} \sqrt{\left(1 + \frac{1}{2}RS^2\right)/T}\right)$.

For example, for 60 observations, if the true value of the Sharpe ratio is equal to 1.50, then the standard deviation of the Sharpe ratio estimator is equal to 0.188. If the true value of the Sharpe ratio is equal to 3, then the standard deviation of the Sharpe ratio estimator is equal to 0.303. Thus, for instance, the hedge funds, which search for high Sharpe ratios, often have a higher standard deviation of the Sharpe ratio estimator than usual funds.

TABLE 5.1: Asymptotic standard deviation of the Sharpe ratio estimator

	Number T of observations					
Sharpe ratio	12	24	36	48	60	120
0.5	0.306	0.217	0.177	0.153	0.137	0.097
0.75	0.327	0.231	0.189	0.163	0.146	0.103
1	0.354	0.250	0.204	0.177	0.158	0.112
1.25	0.385	0.272	0.222	0.193	0.172	0.122
1.5	0.421	0.298	0.243	0.210	0.188	0.133
1.75	0.459	0.325	0.265	0.230	0.205	0.145
2	0.500	0.354	0.289	0.250	0.224	0.158
2.25	0.542	0.384	0.313	0.271	0.243	0.172
2.5	0.586	0.415	0.339	0.293	0.262	0.185
2.75	0.631	0.446	0.364	0.316	0.282	0.200
3	0.677	0.479	0.391	0.339	0.303	0.214

However, most of the time, portfolio returns are not iid, such as hedge funds. Autocorrelations are high. Lo [359] shows that neglecting this feature may lead to overestimation of the Sharpe ratio (65% more for example), and that specific estimators must be used to avoid this problem. □

5.1.3 Other performance measures

5.1.3.1 The information ratio

5.1.3.1.1 Information ratio definition As seen in Chapter 4, many funds are based on benchmark optimization. Thus, such funds must be evaluated by taking account of this feature. Recall that the set of optimal portfolios is another frontier, called the relative frontier. For all efficient portfolios of the standard mean-variance frontier, the Sharpe ratio is constant. For optimal portfolios relative to a benchmark, the new risk measure is usually the tracking error volatility. Then, a new performance measure adjusted to the relative risk can be introduced.

DEFINITION 5.1 *The information ratio is defined by:*

$$RI = \frac{\overline{R}_P - \overline{R}_B}{T}, \tag{5.22}$$

where $T = \sigma\left(R_P - R_B\right)$ is the tracking-error volatility. An ex-post ratio can also be introduced:

$$\widehat{RI} = \frac{\widehat{R}_P - \widehat{R}_B}{\widehat{\sigma}\left(R_P - R_B\right)}. \tag{5.23}$$

PROPOSITION 5.4

As proved in Bertrand et al. [61], all portfolios belonging to the relative frontier (except the benchmark itself) have the same information ratio. Moreover, this ratio does not depend on the benchmark.

PROOF Using relation $T = D\sqrt{\sigma_1^2 - \sigma_0^2}$ (see Chapter 2), we deduce:

$$RI = \frac{(R_P - R_B)}{D\sqrt{\sigma_1^2 - \sigma_0^2}} = \frac{(R_P - R_B)}{\frac{(R_P - R_B)}{(R_1 - R_0)}\sqrt{\sigma_1^2 - \sigma_0^2}} = \frac{R_1 - R_0}{\sqrt{\sigma_1^2 - \sigma_0^2}} = cste. \tag{5.24}$$

Therefore, the information ratio of any portfolio of the relative frontier depends only on characteristics of portfolios 0 and 1 (which belong to the mean-variance frontier).

□

REMARK 5.6 Due to previous properties, the information ratio is adapted to measure performance of benchmark funds. Consider portfolios of iso-beta curves in Chapter 4. Their information ratios depend on beta:

- If $\beta = 1$, portfolios of the iso-beta frontier (except the benchmark), have the same information ratio absolute value.

- If $\beta \neq 1$, portfolios of the iso-beta frontier have different information ratios.

This is a drawback since, for a given beta, the portfolio which has the higher information ratio corresponds to an infinite standard deviation.

On the contrary, portfolios belonging to a same ϕ-frontier (see mean-variance optimization with tracking error constraint in Chapter 4) have the same information ratio, whatever the value of the aversion to variance ϕ. With respect to the relative performance, they are indistinguishable.

□

5.1.3.2 Information ratio and statistical test of excess performance

The information ratio can be viewed as a statistical test of the excess performance w.r.t. the benchmark. Assume that the monthly portfolio and benchmark returns are iid. The excess return is a random variable ER, with expectation μ and unknown standard deviation.

Let $\overline{ER} = \overline{R}_P - \overline{R}_B$ be the empirical mean, and $s = \sigma(R_P - R_B)$ the empirical standard deviation of the excess return, estimated from a sample of n observations. Then, we can consider the following univariate test:

$$\begin{cases} H_0 : \mu \leq 0, \\ H_1 : \mu > 0. \end{cases} \tag{5.25}$$

Under the null hypothesis, $t_c = \sqrt{n}\left(\overline{ER} - 0\right)/s = \sqrt{n}.RI$. The t-statistic $t = \frac{\overline{ER}-\mu}{s/\sqrt{n}}$ has a $(n-1)-$Student distribution. Therefore, the information ratio corresponds to the statistical test (divided by the square root of the number of observations) of the null hypothesis, which corresponds to the case of no overperformance of the portfolio w.r.t. the benchmark.

Assume, for example, that n is equal to 60 (5 years of monthly data). Then, to reject hypothesis H_0 for the threshold 5%, it is necessary that $t_c > 1.671$. The monthly information ratio must be higher than 0.2157 (which corresponds to 0.747 for one year).

Suppose also that this information ratio corresponds to a mean ratio with management cost equal to 0.36. In order to reject hypothesis $\mu \leq 0$ and to conclude that the fund is performant with a confidence level equal to 95%, we must have $\frac{0.36}{\sqrt{12}}\sqrt{n} = 1.645$. This corresponds to $n = 250.5$ months, or equivalently to about 21 years. Thus, this result shows that a long period of performance persistence is needed to conclude that the better performance w.r.t. the benchmark is significant.

5.1.3.3 Adjusted measures

5.1.3.3.1 The Sortino ratio Indicators based on standard deviations do not indicate whether the value is above or below the mean. Risk-adjusted measures based, for example, on semi-variance can separate these two cases. Moreover, they can better take account of asymmetrical return distributions. This is the purpose of the Sortino ratio, introduced in Sortino and Price [477]. It is defined similar to the Sharpe ratio, but with a minimum acceptable return (MAR) instead of the riskless return. Moreover, the standard deviation is replaced by the semi-standard deviation of the return below the MAR.

$$Sortino\ ratio_P = \frac{\overline{R}_P - MAR}{\sqrt{\frac{1}{T}\sum_{t=0}^{T}\left([R_{P_t} - MAR]^+\right)^2}}. \tag{5.26}$$

5.1.3.3.2 The Morningstar rating system The Morningstar rating is also called the risk-adjusted rating (RAR). It is used in particular in the USA. It is based on a system of stars, and is devoted to the ranking of funds belonging to the same peer group. It is calculated as the difference between relative returns and relative risks:

$$RAR_P = RR_P - RRisk_P. \tag{5.27}$$

Relative return and risk are defined as follows:

$$RR_P = \frac{\overline{R}_P}{BR_G},\ RRisk_P = \frac{Risk_P}{BRisk_G}, \tag{5.28}$$

where P is the portfolio belonging to the peer group G, such as domestic stock funds, international stock funds, taxable bond funds, and tax-exempt municipal bond funds as well as equity funds classified by style: capitalization (large-cap, mid-cap, small cap) or growth/value.

$\overline{R}_P$ is the return of fund P in excess of the risk-free rate, and $Risk_P$ is the risk of fund P.

BR_G denotes the base which is used to calculate the relative returns of funds in the group G, and $BRisk_G$ is the base which is used to calculate the relative risks of funds in the group.

The risk of the fund can be measured by first determining the average of the negative values of the fund's monthly returns in excess of the short-term riskless rate, then by setting:

$$Risk_P = -\frac{1}{T}\sum_{t=0}^{T} \min\left[R_{P_t}, 0\right], \tag{5.29}$$

where T is the number of months.

Risk can also be measured by taking account of both downside risk and upside volatility, as well as downward volatility. Then, such criterion penalizes funds with highly volatile returns (both upside and downside). Indeed, extreme gains (upside volatility) may be associated to potential extreme losses (such as Internet funds in the early 2000s). Thus, this kind of performance measure reduces the potential interest in high-risk funds.

The base is calculated by taking the average return of the n funds contained in the group. It can be compared with the riskless rate R_f:

$$BR_G = \max\left(\frac{1}{n}\sum_{i=1}^{n} R_{P_i}, R_f\right). \tag{5.30}$$

The base for the relative risk is defined as the average of the funds in the peer group:

$$BRisk_G = \frac{1}{n}\sum_{i=1}^{n} Risk_{P_i}. \tag{5.31}$$

Sharpe [467] examines properties and limitations of this measure, which is not appropriate for measuring the risk of funds held over a long period.

5.1.3.3.3 The Dowd ratio The performance analysis can be also based on other risk measures, such as the Value-at-Risk measure (see Chapter 2). For example, we can consider a Sharpe-type ratio in which the standard deviation is replaced with a risk measured based on the VaR:

$$\frac{\overline{R}_P - R_f}{VaR_P/V_{P,0}}, \tag{5.32}$$

where VaR_P denotes the VaR of portfolio P, and $V_{P,0}$ is the initial value of this portfolio.

Dowd [170] introduces a measure of investment process based on the VaR.

Consider an investor who holds a portfolio which can be modified by introducing a new asset. To analyze the potential benefit to introduce this asset in the portfolio, the investor defines the risk in terms of the increase in the portfolio's VaR. The asset will be included if the excess return that it gives is not compensated by a too high incremental VaR.

Assuming for example that portfolio returns are normally distributed, the VaR of the portfolio P is proportional to the standard deviation:

$$VaR = -\alpha\sigma(R_P)N, \tag{5.33}$$

where α denotes the confidence parameter for which the VaR is estimated, $\sigma(R_P)$ is the standard deviation of the portfolio returns, and N is a parameter linked to the size of the portfolio. Then, asset A will be included in the portfolio with weight a if and only if:

$$\overline{R}_A \geq \overline{R}_P + \frac{\overline{R}_P}{a}\left(\frac{VaR_{P_A}}{VaR_P} - 1\right), \tag{5.34}$$

where P_A denotes the new portfolio with asset A. Denote the incremental VaR by:

$$IVaR = VaR_{P_A} - VaR_P.$$

Define the function γ_A by:

$$\gamma_A(VaR) = \frac{1}{a}\frac{IVaR}{VaR_P}.$$

The function γ_A is the percentage increase in the VaR due to the purchase of asset A. Then, asset A will be included in the portfolio with weight a if and only if:

$$\overline{R}_A \geq (1 + \gamma_A(VaR))\,\overline{R}_P. \tag{5.35}$$

5.1.4 Beyond the CAPM

Criticisms of the CAPM have led to the introduction of more sophisticated statistical models, such as Arch models. The volatility may depend on past asset values, or the coefficient beta may evolve according to given factors.

5.1.4.1 Performance measurement using a conditional beta

Amenc and Lesourd [22] propose to introduce ARCH modelling for identified factors which influence asset return dynamics (not for the assets themselves, since the number of assets is much higher than the number of factors).

Denote by n the number of assets and by R_t the vector of the returns at time t. Consider the conditional expectation $\mathbb{E}\left[R_t\,|\mathcal{F}_{t-1}\right]$ and variance-covariance matrix $\mathbb{V}\left[R_t\,|\mathcal{F}_{t-1}\right]$ of the vector of returns. Assume that these terms are given by the following equation:

$$R_t = \beta f_t + u_t, \tag{5.36}$$

with

$$\begin{aligned} \mathbb{E}\left[u_t\,|\mathcal{F}_{t-1}\right] &= 0, \\ \mathbb{V}\left[R_t\,|\mathcal{F}_{t-1}\right] &= \sigma^2. \end{aligned} \tag{5.37}$$

The factor f_t satisfies

$$f_t = \mu + \varepsilon_t, \tag{5.38}$$

with $(\varepsilon_t)_t$ and $(u_t)_t$ independent and

$$\begin{aligned} &\mathbb{E}\left[\varepsilon_t\,|\mathcal{F}_{t-1}\right] = 0, \\ &\mathbb{V}\left[\varepsilon_t\,|\mathcal{F}_{t-1}\right] = c + \sum_{i=1}^{q} a_i \varepsilon_{t-1}^2 + \sum_{j=1}^{p} b_j h_{t-j}. \end{aligned} \tag{5.39}$$

Thus

$$\mathbb{E}\left[R_t\,|\mathcal{F}_{t-1}\right] = \delta\beta \text{ where } \delta \text{ is a scalar.}$$

In addition, the variance-covariance matrix of the n assets can be written as follows:

$$\mathbb{V}\left[R_t\,|\mathcal{F}_{t-1}\right] = \begin{pmatrix} \beta_1^2 ... \beta_1\beta_n \\ ... \\ \beta_1\beta_n ... \beta_n^2 \end{pmatrix} \mathbb{V}\left[\varepsilon_t\,|\mathcal{F}_{t-1}\right] + \begin{pmatrix} {\sigma_1}^2 \; 0...0 \\ 0... \\ 0...0 \; \beta \; \sigma_n^2 \end{pmatrix}. \tag{5.40}$$

Therefore, for a portfolio with weights $(w_1, ..., w_n)$, we have:

$$R_{P_t} = \sum_{i=1}^{n} w_{i,t} R_{i,t} = \sum_{i=1}^{n} w_{i,t} \left(\beta_i f_{i,t} + u_{i,t}\right). \tag{5.41}$$

Set $\beta_{P,t} = \sum_{i=1}^{n} w_{i,t}\beta_i$ and $u_{P,t} = \sum_{i=1}^{n} w_{i,t}u_{i,t}$. Then, the Jensen measure is estimated from the relation

$$R_{P_t} = \alpha_P + \beta_{P,t} f_t + u_{P,t}. \tag{5.42}$$

The coefficients α_P and $\beta_{P,t}$ are estimated by using regression from a time series on the portfolio and the market.

5.1.4.2 Performance measurement using a conditional beta

A conditional version of the CAPM can be introduced, as in Ferson and Schadt [222]. The conditional CAPM is such that, for any asset i:

$$R_{i,_{t+1}} - R_{f,t} = \beta_{i,M,t}\ \left(R_{M,_{t+1}} - R_{f,t}\right) + u_{i,t+1}, \tag{5.43}$$

with

$$\begin{aligned} \mathbb{E}\left[u_{i,t+1}\left|\mathcal{F}_t\right.\right] &= 0, \\ \mathbb{E}\left[u_{i,t+1}\left(R_{M,_{t+1}} - R_{f,t}\right)\left|\mathcal{F}_t\right.\right] &= 0. \end{aligned} \tag{5.44}$$

The filtration $\mathcal{F}_t$ represents the public information available at time t. The beta $\beta_{i,M,t}$ of the regression is a conditional beta which depends on the information $\mathcal{F}_t$, generated by a given set of factors.

If only information $\mathcal{F}_t$ is used, then the alpha term is null. The error term $u_{i,t+1}$ in the regression is independent from information $\mathcal{F}_t$. Therefore, this model is coherent with the efficient market hypothesis. If $\mathcal{F}_t$ is the $\sigma-$algebra generated by a process $(I_t)_t$, then, using Taylor's approximation, the beta can be estimated through a linear function:

$$\beta_{P,M,t} = b_{0,P} + B'_P(I_t - \overline{I}_t), \tag{5.45}$$

where $b_{0,P}$ is an average beta which is equal to the unconditional mean of the conditional beta: $b_{0,P} = \mathbb{E}\left[\beta_{P,M,t}\left|\mathcal{F}_t\right.\right]$. The components of the vector B_P are the response coefficients of the conditional beta w.r.t. the random variable I_t. Then, a conditional formulation of portfolio return can be deduced:

$$R_{P,_{t+1}} - R_{f,t+1} = b_{0,P}\left(R_{M,_{t+1}} - R_{f,t+1}\right) + B'_P\ I_t\left(R_{M,_{t+1}} - R_{f,t+1}\right) + u_{P,t+1}, \tag{5.46}$$

with

$$\begin{aligned} \mathbb{E}\left[u_{P,t+1}\left|\mathcal{F}_t\right.\right] &= 0, \\ \mathbb{E}\left[u_{P,t+1}\left(R_{M,_{t+1}} - R_{f,t+1}\right)\left|\mathcal{F}_t\right.\right] &= 0. \end{aligned} \tag{5.47}$$

Thus, this is a stochastic factor model which is a linear function of the excess market return, the coefficients of which depend linearly on I_t. It is

a time-dependent generalization of the standard CAPM and can be used to examine, for example, the Jensen measure:

$$R_{P,_{t+1}} - R_{f,t+1} =$$

$$\alpha_{CP} + b_{0,P}\left(R_{M,_{t+1}} - R_{f,t+1}\right) + B'_P\ I_t\left(R_{M,_{t+1}} - R_{f,t+1}\right) + e_{P,t+1}, \quad (5.48)$$

where α_{CP} is the average difference between the excess return of the portfolio and the excess return of a dynamic reference strategy. To improve the alpha forecast, Ferson and Schadt [222] assume a relationship between the portfolio risk and the market indicators: for example, the market index dividend yield $(DY_t)_t$ and the return on short-term T-bills $(TB_t)_t$.

Denote dy_t and tb_t the random variables equal to the differentials compared to the average of the variables DY_t and TB_t:

$$\begin{aligned} dy_t &= DY_t - \mathbb{E}\left[DY_t\right], \\ tb_t &= TB_t - \mathbb{E}\left[TB_t\right]. \end{aligned} \quad (5.49)$$

Therefore, we have: $I_t - \overline{I}_t = \begin{bmatrix} dy_t \\ tb_t \end{bmatrix}$ and $B_P = \begin{bmatrix} b_1 P \\ b_{2P} \end{bmatrix}$. The conditional beta can be written as:

$$\beta P = b_0 + b_1 dy_t + b_2 tb_t.$$

Then, the conditional formulation of the Jensen model is given by:

$$\begin{aligned} R_{P,_{t+1}} - R_{f,t+1} = {} & \alpha_{CP} + b_{0,P}\left(R_{M,_{t+1}} - R_{f,t+1}\right) && (5.50) \\ & + \left(b_1 dy_t + b_2 tb_t\right)\left(R_{M,_{t+1}} - R_{f,t+1}\right) + e_{P,t+1}, && (5.51) \end{aligned}$$

where α_{CP} is the conditional performance measure, $b_{0,P}$ is the conditional beta, and b_1 and b_2 indicate the variations in the conditional beta compared with the dividend yield and the return on the T-bills. These coefficients are estimated through regression from the time series of the variables.

5.1.4.3 Performance measurement independent of the market model

Due to Roll's criticism, measures that do not depend on the market model have been developed, such as the Cornell measure and the Grinblatt and Titman measure. The goal of these measures is to evaluate the managers's capacity to select stocks that have higher returns than the average. This average must be defined, which is one drawback of these measures.

5.1.4.3.1 The Cornell measure The Cornell measure (see Cornell [130]) is defined as the average difference between the return on the investor's portfolio during the management period, and the return of the reference portfolio

with the same weighting but observed on a different period than the investor's management period. The calculation is made after the investment horizon.

The Cornell measure can be written as:

$$C = \widehat{R}_P - \widehat{\beta}_P \widehat{R}_B, \tag{5.52}$$

where the symbol $\widehat{X}_P$ denotes the $\mathbb{P}$-limit of $\frac{1}{T}\sum_{t=1}^{T} X_{P,t}$.

Therefore, we get:

$$C = \mathbb{P} - \text{limit of } \left[\frac{1}{T}\sum_{t=1}^{T} \beta_{P,t}\left(R_{B,t} - \widehat{R}_{B,t}\right)\right] + \widehat{\varepsilon}_P. \tag{5.53}$$

This means that C is the sum of the selectivity and timing components in the Jensen measure decomposition. Thus, if the investor has no particular skill in terms of timing or selectivity, Jensen and Cornell measures give a null performance.

5.1.4.3.2 The Grinblatt and Titman measures Grinblatt and Titman [269], [270]) propose a measure to improve upon the Jensen measure by allowing us to better take account of market timing, without any information about portfolio weighting. This method assumes that, if a manager has a market timing skill, then his performance would be observed over several periods. The measure attributes a null performance to uninformed fund managers. It is defined as follows. Let $w_{P,t}$ denote the weighting attributed to the return for period t. Consider the return $R_{B,t}$ of the reference portfolio for the period t. We have: $\sum_{t=1}^{T} w_{P,t}\left(R_{B,t} - R_{F,t}\right) = 0$. Then, a first Grinblatt and Titman measure is given by:

$$GB1 = \sum_{t=1}^{T} w_{P,t}\left(R_{P,t} - R_{f,t}\right). \tag{5.54}$$

Therefore, a positive GB indicates that the fund manager has good forecasts on market dynamics. However, this measure supposes that the portfolio weighting is well determined.

For this reason, Grinblatt and Titman [271] introduce another measure which takes account of the portfolio's composition evolution. This approach supposes that an informed fund manager modifies the portfolio weights according to his forecast on the market's evolution. Thus, covariances between asset returns and their weights are not null. The measure involves these covariances:

$$GB2 = \frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(w_{i,t} - w_{i,t-k}\right)\left(R_{i,t} - R_{f,t}\right), \tag{5.55}$$

where $w_{i,t}$ and $w_{i,t-k}$ are the weights of asset i at times t and $t-k$.

The expectation of this measure is null if the fund manager is not informed, and positive otherwise. No reference portfolio is used, but this method requires a large amount of data and calculation.

5.1.4.4 Performance measurement and multi-factor models

Generalizations of the CAPM can be based on multi-factor models, which are also linear models but do not make any assumption on the investor's risk aversion.

A first model, introduced by Ross [439], is based on a specific arbitrage valuation. It is called the Arbitrage Pricing Theory (APT). It does not assume normality of returns and supposes only that investors are risk-averse, without specifying a particular utility function. We have:

$$R_{i,t} = \mathbb{E}\left[R_i\right] + \sum_{k=1}^{K} b_{i,k} F_{k,t} + \varepsilon_{i,t}, \tag{5.56}$$

where $b_{i,k}$ denotes the sensitivity of asset i to factor k, $F_{k,t}$ denotes the return of factor k with $\mathbb{E}\left[F_{k,t}\right] = 0$, and $\varepsilon_{i,t}$ denotes the residual return of asset i. It is the specific risk of asset i which is not explained by the factors and satisfies:

$$\mathbb{E}\left[\varepsilon_{i,t}\right] = 0.$$

- The APT model supposes that markets are perfectly efficient and that the factor model is the same for all investors.
- The number n of assets is assumed to be very large w.r.t. the number K of factors.

The residuals are independent from each other and independent from the factors:

$$\begin{aligned} Cov(\varepsilon_{i,t}, \varepsilon_{j,t}) &= 0, \forall i \neq j, \\ Cov(\varepsilon_{i,t}, F_{k,t}) &= 0, \forall i, \forall k. \end{aligned} \tag{5.57}$$

Arbitrage conditions lead to the existence of factor risk premia λ_k such that:

$$\mathbb{E}\left[R_i\right] - R_f = \sum_{k=1}^{K} \lambda_k b_{i,k}. \tag{5.58}$$

Denote δ_k the expected return of a portfolio with a sensitivity to factor k equal to 1, and null sensitivity to other factors. Then:

$$\lambda_k = \delta_k - R_f \text{ and } \mathbb{E}\left[R_i\right] - R_f = \sum_{k=1}^{K} \left(\delta_k - R_f\right) b_{i,k}, \tag{5.59}$$

where $b_{i,k} = \frac{Cov(R_i, \delta_k)}{Var(\delta_k)}$ are the sensitivities to the factor loadings.

When there exists only one factor corresponding to the market return, this model is the CAPM. The problem is to identify the number of factors. Numerous empirical studies are devoted to the determination of the macroeconomic or financial factors. For example, the three-factor model of Fama and French [219] takes account of the *book-to-market* ratio and the company's size measured by its market capitalization:

$$R_{Pt} - R_{ft} = \alpha_P + \beta_P \left(R_{Mt} - R_{ft}\right) + b_S.SMB_t + b_H.HML_t + \varepsilon_{Pt}, \quad (5.60)$$

where SMB_t indicates "*small (cap) minus big.*" It measures the excess return of the small-capitalization returns w.r.t. large-capitalization returns.

The term HML_t is the "*high (book/price) minus low.*" It denotes the difference between returns on portfolios with high book-to-market ratios and portfolios with low book-to-market ratios.

Fama and French assume that the market is efficient, but that more than one factor is needed to explain asset returns.

Carhart's four factor model [105] introduces one additional factor: the $PRIYR$ which denotes the difference between the average of the highest returns and the average of the lowest returns from the previous year. Thus, we have the following decomposition:

$$R_{Pt} - R_{ft} = \alpha_P + \beta_P \left(R_{Mt} - R_{ft}\right) + b_S.SMB_t + b_H.HML_t + b_P.PRIYR_t + \varepsilon_{Pt} \quad (5.61)$$

The Barra multi-factor model (see Barra [42] and Scheikh [453]) supposes that asset returns are determined by the firm's characteristics: size, earnings, industrial sectors, *etc.*, which are the factors used jointly with risk indices.

REMARK 5.7 As can be seen, the application of multi-factor model to performance measurement requires the choice of a specific model to take account of industrial and financial factors. Endogeneous factors can also be extracted through a factor analysis, either based on the principle of the maximum likelihood, or based on principal component analysis. It also provides for factor loadings. However, in that case, these factors are not well-identified. Therefore, we can search for explanation of these factors from known indicators. Then, we can deduce an explicit decomposition from the implicit decomposition.

$\square$

5.2 Performance decomposition

Since nowadays investors request more information about the investment process of fund managers, alternative performance measures are necessary to search for the performance causes. The performance attribution tries to decompose the excess performance into identified terms by taking more account of the management process.

5.2.1 The Fama decomposition

Fama [218] proposes a performance decomposition which separates the fund performance into two parts:

- The selectivity

  and

- The risk.

This analysis is made in the CAPM framework. A portfolio P is compared with a “naïve” portfolio C, which consists of investing a weight x on the riskless asset, and $(1-x)$ on the market portfolio, such that the portfolio beta of C is equal to the beta of $P : \beta_C = \beta_P$. Then, the portfolio C is the efficient portfolio which has the same systematic risk as portfolio P and does not require a forecast ability. Then, we have:

$$\bar{R}_C = (1-\beta_P)\, R_f + \beta_P \bar{R}_M. \tag{5.62}$$

The portfolio P may be less diversified. This risk may allow for an excess performance w.r.t. portfolio C.

Fama proposes the following decomposition:

$$\overbrace{\bar{R}_P - R_f}^{\text{Total performance}} = \overbrace{\left[\bar{R}_P - \bar{R}_C\right]}^{\text{Selectivity}} + \overbrace{\left[\bar{R}_C - \bar{R}_f\right]}^{\text{Risk}}. \tag{5.63}$$

The selectivity term measures the performance part due to the systematic risk, born by the fund manager. It is equal to the Jensenalpha. Portfolios P and C have the same systematic risk. However, their total risks are different. If portfolio P is the main investor's wealth, then the total risk is the pertinent risk measure. In that case, portfolio P must be compared with a naïve portfolio with the same total risk rather than the same sytematic risk.

Thus, Fama proposes the following selectivity decomposition:

$$\overbrace{\left[\bar{R}_P - \bar{R}_C\right]}^{\text{Selectivity}} = \text{Net Selectivity} + \overbrace{\left[\bar{R}_{C'} - \bar{R}_C\right]}^{\text{Diversification}}, \tag{5.64}$$

where

$$\text{Net Selectivity} = \overbrace{\left[\bar{R}_P - \bar{R}_C\right]}^{\text{Selectivity}} - \overbrace{\left[\bar{R}_{C'} - \bar{R}_C\right]}^{\text{Diversification}}. \quad (5.65)$$

By definition, $R_{C'}$ is the return of a portfolio which contains the riskless asset and the market portfolio with the same total risk as portfolio P. The portfolio C' is also efficient. If the net selectivity is negative, then the fund manager has a diversified risk which has not been compensated by excess return. The diversification term takes account of this additional return (*i.e.*, $\sigma(R_P) - \beta_P$). It is always positive.

The risk $\bar{R}_C - R_f$ can also be decomposed as follows:

$$\overbrace{\left[\bar{R}_C - R_f\right]}^{\text{Risk}} = \overbrace{\bar{R}_C - \bar{R}_T}^{\text{Manager's risk}} + \overbrace{\left[\bar{R}_T - R_f\right]}^{\text{Investor's risk}}. \quad (5.66)$$

The investor has a risk objective β_T which leads to a portfolio with return $\bar{R}_P$ on the security market line. The fund manager chooses a portfolio with risk β_P. Therefore, the performance component due to the total risk is due to the risk level fixed by the investor, and to the risk level fixed by the fund manager.

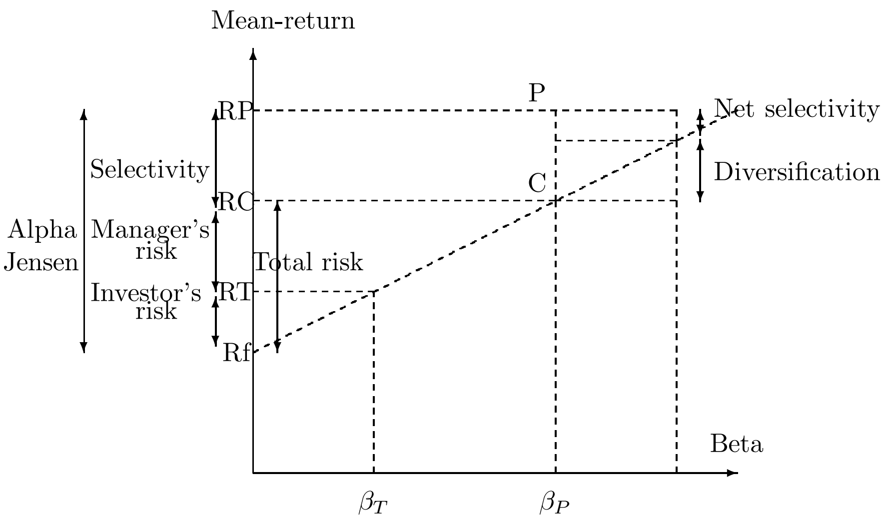


FIGURE 5.5: FAMA performance decomposition

5.2.2 Other performance attributions

Two kinds of performance attribution can be distinguished:

- The external attribution, which uses exogeneous information w.r.t. the time series of the portfolio and benchmark returns.
- The internal attribution, which uses the time series of the portfolio and benchmark weightings.

5.2.3 The external attribution

The following methods allow for the determination of the fund manager's ability to forecast the global evolution of the market and to select assets well.

5.2.3.1 Treynor-Mazuy

In order to determine the quality of the forecast concerning the global evolution of the market, Treynor and Mazuy[495] examine the beta evolution w.r.t. market evolution. Typically, a successful *market timing* strategy is associated to a beta which is higher than 1 when the market is bullish, and smaller than 1 when the market is bearish. If the fund manager has no market timing goal, then $\beta_P = cste$. Then, without specific risk, points with components (R_{Mt}, R_{Pt}) would be on the straight line with equation: $R_{Pt} = \beta_P R_{Mt}$. With specific risk, but still with $\beta_P = cste$, one would observe a set of points such as in the following figure.

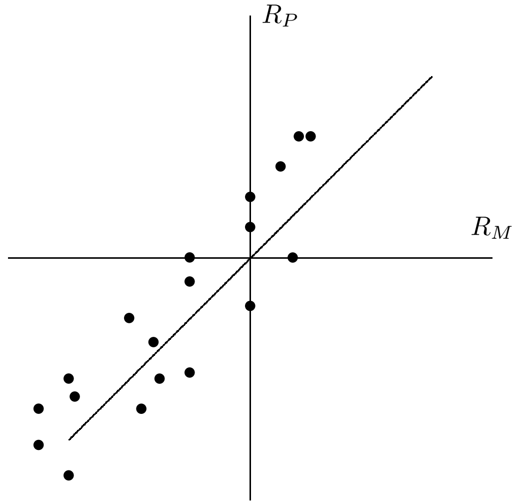


FIGURE 5.6: Linear regression of R_P on R_M without market timing

However, as soon as the fund manager tries to forecast the market evolution and modifies his beta according to this, one would observe points in the previous figure above the regression line for various market return values R_{Mt}. The set of points would be adjusted w.r.t. a convex curve, if the market timing strategy is successful, as shown in Figure (5.7).

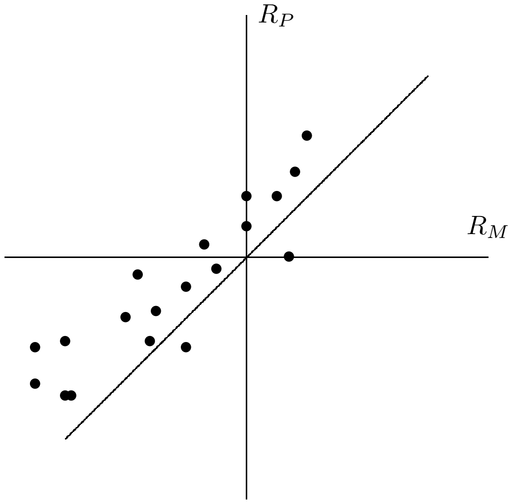


FIGURE 5.7: Linear regression of R_P on R_M with successful market timing

Treynor and Mazuy propose a statistical method based on this property. More precisely, in order to allow for a convex relation between R_{Mt} and R_{Pt}, they introduce the following equation:

$$R_{Pt} - R_{ft} = \alpha_P + \beta_P \left(R_{Mt} - R_{ft}\right) + \delta_P \left(R_{Mt} - R_{ft}\right)^2 + \varepsilon_{Pt}. \tag{5.67}$$

If coefficient δ_P is significantly different from zero, then we can conclude that the fund manager has used a market-timing strategy (with success if $\delta_P > 0$ and failure if $\delta_P < 0$).

Moreover, the alpha coefficient is still viewed as the fund manager's ability to select the assets with returns above the values given by the CAPM.

5.2.3.2 Henrickson-Merton

Henrickson and Merton [290] assume that the fund manager can forecast a comparison between R_{Mt} and R_{ft}: $R_{Mt} > R_{ft}$ or $R_{Mt} < R_{ft}$. This assumption leads to the following regression model:

$$R_{Pt} - R_{ft} = \alpha_P + \beta_P (R_{Mt} - R_{ft}) + \delta_P D (R_{Mt} - R_{ft}) + \varepsilon_{Pt}, \qquad (5.68)$$

with $D = 1$ if $R_{Mt} > R_{ft}$, and $D = 0$ if $R_{Mt} \leq R_{ft}$.
Then, the beta can take two values: $\beta_P + \delta_P$ (resp. β_P) is the portfolio beta when the market return is higher (resp. smaller) than the riskless return. A market-timing strategy is successful if δ_P is significantly positive. The fund manager's ability to select assets is determined from the Jensen alpha.

5.2.4 The internal attribution

The purpose of this performance attribution is to explain the excess performance w.r.t. a benchmark, jointly determined by the investor and the fund manager. As seen in Chapter 4, the management process can be divided into the asset allocations and the tactical strategy based on stock picking and market timing.

5.2.4.1 The Brinson model

Brinson *et al.* [90] propose the following method: consider a portfolio P and a benchmark B which contain n asset classes $i = 1, ..., n$.

Denote:

- U the set of assets, $l = 1, ..., m = \#(U)$, $\{U_1, ..., U_n\}$ is a partition of the set U and is the set of the n asset classes $i = 1, ..., n \leq m$;
- Z_l the return of asset l (vector: Z). This is a random variable if we consider ex-ante values. This is a scalar if ex-post data are analyzed;
- R_l the expected return of asset l (vector : R);
- σ_{kl} the covariance between assets k and l (matrix : V);
- z_{pl} (*resp.* z_{bl}) the weight on asset l in the portfolio (*resp.* the *benchmark*) (vector: z_p and z_b);
- $w_{pi} = \sum\limits_{l \in U_i} z_{pl}$ (*resp.* $w_{bi} = \sum\limits_{l \in U_i} z_{bl}$) the weight of asset class i in the portfolio (*resp.* the benchmark);
- $\varpi_{pl} = \frac{z_{pl}}{\sum\limits_{l \in U_i} z_{pl}}$ and $\varpi_{bl} = \frac{z_{bl}}{\sum\limits_{l \in U_i} z_{bl}}$ are the weights of asset l into the asset class i of the portfolio and the benchmark;
- $Z_{pi} = \sum\limits_{l \in U_i} \varpi_{pl}.Z_l$ (*resp.* $Z_{bi} = \sum\limits_{l \in U_i} \varpi_{bl}.Z_l$) is the return of asset class i in the portfolio (*resp.* the benchmark); and,
- $R_{pi} = \sum\limits_{l \in U_i} \varpi_{pl}.R_l$ (*resp.* $R_{bi} = \sum\limits_{l \in U_i} \varpi_{bl}.R_l$) is the mean return of asset class i in the portfolio (*resp.* in the benchmark).

Then, the excess global performance of the portfolio w.r.t. the benchmark is given by:

$$S = R_p - R_b = \sum_{i=1}^{n} (w_{pi}.R_{pi} - w_{bi}.R_{bi}) = \sum_{l=1}^{m} (z_{pl} - z_{bl}) R_l. \tag{5.69}$$

This excess performance has to be decomposed into main management processes: asset allocation and asset selection.

5.2.4.1.1 Asset allocation effect When searching for the origin of excess performance, we encounter a problem due to the allocation effect for one particular asset class. Indeed, the high weighting (resp. the low weighting) of an asset class leads to the low or high weighting of at least one another class.

One would expect to consider the following measure for asset class i:

$$(w_{pi} - w_{bi})\, R_{bi}. \tag{5.70}$$

However the following example shows that this measure is not well adapted.

Example 5.2

TABLE 5.2: Asset allocation (percentages)

Class	w_{pi}	w_{bi}	R_{bi}	$(w_{pi} - w_{bi})R_{bi}$
1	25	15	8	0.8
2	40	45	10	−0.5
3	35	40	12	−0.6
Total	100	100	$R_b = 10.50$	−0.3

The last column leads to the conclusion that the high weighting of asset class 1 is judicious, and that the relatively bad global performance of the portfolio is due to the low weighting of asset classes 2 and 3. Nevertheless, we can see that the fund manager has highly weighted an asset class which has a return smaller than the mean return of the benchmark (8% versus 10.5%). □

This example proves that we must take into account the difference between the class return and the mean benchmark return. Therefore, the contribution of asset class i to the excess performance is defined by:

$$AA_i = (w_{pi} - w_{bi})\,(R_{bi} - R_b)\,. \tag{5.71}$$

The following table indicates the different cases:

TABLE 5.3: Contribution to asset classes

	Excess performance class i: $(R_{bi} - R_b) > 0$	Bad performance class i: $(R_{bi} - R_b) < 0$
High weighting class i: $(w_{pi} - w_{bi}) > 0$	Good decision $(w_{pi} - w_{bi})\,.$ $(R_{bi} - R_b) > 0$	Bad decision $(w_{pi} - w_{bi})\,.$ $(R_{bi} - R_b) < 0$
Low weighting class i: $(w_{pi} - w_{bi}) < 0$	Bad decision $(w_{pi} - w_{bi})\,.$ $(R_{bi} - R_b) < 0$	Good decision $(w_{pi} - w_{bi})\,.$ $(R_{bi} - R_b) > 0$

Applying this method to the previous example, we get the following table:

TABLE 5.4: Asset selection effects

Class	w_{pi}	w_{bi}	R_{bi}	$(w_{pi} - w_{bi})(R_{bi} - R_b)$
1	25	15	8	−0.25
2	40	45	10	0.03
3	35	40	12	−0.08
Total	100	100	$R_b = 10.50$	−0.30

Note that the bad performance of the portfolio w.r.t. the benchmark is mainly due to the high weighting of asset class 1, which is rather intuitive. Additionally, the global bad performance is the same as previously (−0.3%). This is due to:

$$\sum_{i=1}^{n} (w_{pi} - w_{bi})\, R_b = 0 \text{ since } \sum_{i=1}^{n} w_{pi} = \sum_{i=1}^{n} w_{bi} = 1.$$

5.2.4.1.2 Asset selection effect The contribution to the global excess return of asset selection within each asset class is given by:

$$SE_i = w_{bi}\,(R_{pi} - R_{bi})\,. \tag{5.72}$$

The choice of the benchmark weight for asset class i is justified in order to not interfere with the allocation effect. The difference $(R_{pi} - R_{bi})$ is not null as soon as the fund manager's weighting of the assets included in the asset class i is different from the benchmark's. Therefore, this allows us to measure the selection effect.

5.2.4.1.3 The interaction term Since the sum of the two previous effects is not equal to the global excess performance of asset class i, an additional term must be introduced:

$$I_i = (w_{pi} - w_{bi})\,(R_{pi} - R_{bi})\,. \tag{5.73}$$

Then we deduce:

$$S = \sum_{i=1}^{n} (w_{pi}.R_{pi} - w_{bi}.R_{bi}) = \sum_{i=1}^{n} (AA_i + SE_i + I_i)\,. \tag{5.74}$$

Example 5.3
Consider a fund with benchmark: 35% domestic stocks; 50% domestic bonds; and 15% international stocks. Suppose that the management period corresponds to one year, and that the weighting of asset classes has been determined through a strategical asset allocation which is not modified during the given period. The numerical values are given in the following tables:

TABLE 5.5: Portfolio characteristics

	Portfolio weights	Benchmark weights	Return class i : portfolio	Return class i : benchmark
Class	w_{pi}	w_{bi}	R_{pi}	R_{bi}
Stock (do.)	40	35	13.0	12.0
Bond (do.)	40	50	6.75	7.0
Stock (int.)	20	15	11.0	11.0
Total	100	100		$Rb = 9.35$

The measures of the three effects are given by:

TABLE 5.6: Performance attribution

	Measure of Allocation effect	Measure of selection effect	Measure of interaction effect
Stocks (do.)	0.13	0.35	0.05
Bonds (do.)	0.24	−0.13	0.03
Stocks (Int.)	0.08	0.00	0.00
Total	0.45	0.23	0.075

The benchmark and the portfolio have returns respectively equal to 9.35% and 10.10%. This excess performance is decomposed into 0.45% for the al-

location effect, 0.23% for the asset selection and 0.075% for the interaction effect. □

5.2.4.2 Limit of attribution method

The Brinson method does not sufficiently take into account the risk of financial investments. Consider, for instance, a fund manager who chooses a portfolio according to the Markowitz criterion applied on the tracking error (see Chapter 4 and Roll [431]).

Suppose that the benchmark contains three asset classes, each with two assets. Suppose that this benchmark is not efficient.

The minimization of the tracking error is illustrated by the following figure, where the benchmark (resp. the managed portfolio 1) has a mean-return equal to 8.60% *(resp.* 9.60%) and a standard deviation 9.30% (resp. 10.57%). The tracking error volatility is equal to 2.24%.

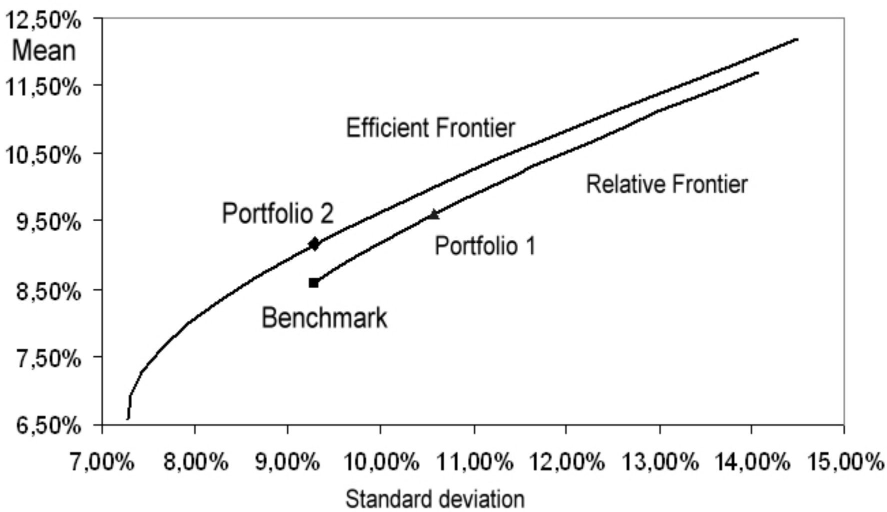


FIGURE 5.8: Efficient and relative frontiers

Despite the optimal choice of portfolio 1, the performance attribution is not necessarily good. The following table indicates the performance attribution process.

TABLE 5.7: Performance attribution of portfolio 1

Asset class	Portfolio weighting	Portfolio weighting	Allocation Effect	Selection Effect	Interaction Effect
1	65.80	40.00	0.232	−0.027	−0.017
2	14.73	40.00	0.278	0.012	−0.007
3	19.47	20.00	−0.002	0.546	−0.014
Total	100.00	100.00	0.508	0.531	−0.039

The excess portfolio return w.r.t. the benchmark is equal to 1.00%: 0.5081% is due to the allocation process, 0.5308% is due to the asset selection and −0.0389% is due to the interaction effect.

Whereas the portfolio is optimal, some of the performance indicators are negative. For example, the low weighting of asset class 3 leads to a negative contribution equal to −0.0002% of the global portfolio performance. However, these results are not too opposed to the portfolio optimality since performance attribution does not take account of the tracking error itself.

5.2.4.2.1 The risk attribution To keep the coherence with the benchmark optimization, Bertrand [56] proposes a performance attribution method based on the decomposition of the tracking-error volatility.

In particular, we can note that information ratios for each decision process are equal (asset allocation, asset selection, and interaction effect).

The tracking-error volatility is given by:

$$
\begin{aligned}
T &= \frac{T^2}{T} = \frac{\left(z_p - z_b\right)' V \left(z_p - z_b\right)}{\sqrt{\left(z_p - z_b\right)' V \left(z_p - z_b\right)}} = \frac{Cov\left(S, S\right)}{T}, \\
&= \frac{1}{T} Cov\left(\sum\nolimits_{l=1}^{m} \left(z_{pl} - z_{bl}\right) Z_l, \sum\nolimits_{l=1}^{m} \left(z_{pl} - z_{bl}\right) Z_l\right), \\
&= \frac{1}{T}\left[Cov\left(\sum_{i=1}^{n} \left(w_{pi} - w_{bi}\right)\left(Z_{bi} - Z_b\right), S\right)\right. \\
&\quad + Cov\left(\sum_{i=1}^{n} w_{bi}\left(Z_{bi} - Z_{bi}\right), S\right) \\
&\quad \left. + \, Cov\left(\sum_{i=1}^{n} \left(w_{pi} - w_{bi}\right)\left(Z_{pi} - Z_{bi}\right), S\right)\right], \\
&= \frac{1}{T}\left[Cov\left(AA, S\right) + Cov\left(SE, S\right) + Cov\left(I, S\right)\right].
\end{aligned}
$$

From the previous relation, the tracking-error volatility T is decomposed into three terms:

- The contribution to the total risk of the asset allocation, which is measured by $Cov\left(AA,S\right)/T$;
- The contribution to the relative risk of the asset selection, which is measured by $Cov\left(SE,S\right)/T$;
- The contribution to the relative risk of the interaction effect, which is measured by $Cov\left(I,S\right)/T$.

PROPOSITION 5.5

According to Bertrand [56]), for any portfolio belonging to the relative frontier, we have:

- *Each term in the risk attribution decomposition has the same sign as the corresponding term in the perfomance attribution decomposition. Additionally, it is equal to the component of performance attribution divided by the information ratio:*

$$\begin{aligned}\frac{Cov\left(AA_i,S\right)}{T} &= \frac{\left(w_{pi}-w_{bi}\right)\left(R_{bi}-R_b\right)}{RI_P},\\ \frac{Cov\left(SE_i,S\right)}{T} &= \frac{w_{bi}\left(R_{bi}-R_{bi}\right)}{RI_P},\\ \frac{Cov\left(I_i,S\right)}{T} &= \frac{\left(w_{pi}-w_{bi}\right)\left(R_{pi}-R_{bi}\right)}{RI_P}.\end{aligned}$$

- *Each term in the risk attribution decomposition has the same information ratio as the information ratio of the portfolio RI_P:*

$$\begin{aligned}RI\left(AA_i\right) &= \frac{\left(w_{pi}-w_{bi}\right)\left(R_{bi}-R_b\right)}{\frac{Cov\left(AA_i,S\right)}{T}} = RI_P,\\ RI\left(SE_i\right) &= \frac{w_{bi}\left(R_{bi}-R_{bi}\right)}{\frac{Cov\left(SE_i,S\right)}{T}} = RI_P,\\ RI\left(I_i\right) &= \frac{\left(w_{pi}-w_{bi}\right)\left(R_{pi}-R_{bi}\right)}{\frac{Cov\left(I_i,S\right)}{T}} = RI_P.\end{aligned}$$

The previous results show that for portfolios belonging to the relative frontier, the appropriate risk attribution measure is the tracking-error volatility of each term of the decomposition.

The following table indicates the tracking-error volatility of each component of the performance attribution process proposed in the previous example.

TABLE 5.8: Tracking-error volatilities

	$Cov\,(AA_i, S)$	$Cov\,(SE_i, S)$	$Cov\,(I_i, S)$	Total
Asset class 1	0.5205	−0.0596	−0.0385	0.422
Asset class 2	0.6231	0.0258	−0.0163	0.633
Asset class 3	−0.0047	1.2234	−0.0323	1.186
Total	1.1388	1.1896	−0.0871	2.241

The total tracking-error volatility of the portfolio is equal to 2.241%. Note that each decision which leads to a bad performance (for example, the low weighting of asset class 3 or the weighting into asset class 1) is now clearly identified as a decision which contributes to the reduction of the relative risk and thus is justified.

The constance of information ratio is illustrated by the following table:

TABLE 5.9: Information ratios

	$RI\,(AA_i)$	$RI\,(SE_i)$	$RI\,(I_i)$
Asset class 1	0.44617	0.44617	0.44617
Asset class 2	0.44617	0.44617	0.44617
Asset class 3	0.44617	0.44617	0.44617

Note that Grinold and Kahn [273] consider that such a value for an information ratio (=0.44617) is a good indicator for the active fund manager.

5.3 Further Reading

General results about portfolio performance are presented in Grinold and Kahn [273], and Amenc and Le Sourd [22].

Performance measures taking management style into account are introduced in Sharpe [465], and in Lobosco [360], who proposes the SRAP measure (*style/risk-adjusted performance*). Muralidhar [396] introduces a specific measure to compare performance of different managers within funds with the same objectives (so belonging to the same peer group). International diversification can be taken into account by using the IAPM (*international asset pricing model*) introduced by Solnik [472].

The problem of performance persistence is related to market efficiency. However, from the professional point of view, investment performances for individual fund managers are examined (skillful or lucky?) rather than the global market inefficiency. As mentioned in Kahn and Rudd [313], the earliest empirical studies seem to suggest no performance persistence while recent articles conclude that a certain level of performance persistence exists.

According to Brown *et al.* [93], short-term performance is persistent, but the survivorship bias affects the results (*i.e.*, bad funds tend to disappear), since it is in favor with performance persistence (see also [93]). Jegadesh and Titman [299] examine HYSE and AMEX securities over the period 1965-1989. They show that a momentum strategy, which is based on buying the best funds and selling the worst ones from the previous six months, provides a 1% per month excess return over the following six months.

The difference in results concerning the performance persistence can be due to seasonal or daily effects. Note also that stock markets are subject to cycles. Thus, a given management process may be good for a given cycle and bad for another.

Other recent methods involving new risk measures have been introduced to study risk attribution and portfolio performance. The total risk of a portfolio can be decomposed into terms that can be interpreted as the risk contribution of the corresponding subsets of the portfolio. An overview of such methodology is provided in Rachev and Zhang [422].

Part III

Dynamic portfolio optimization

"Finance is a highly analytical subject, and nowhere more so than in continuous-time analysis. Indeed, the mathematics of the continuous-time finance model contains some of the most beautiful applications of probability and optimization theory. But, of course, not all that is beautiful in science need also be practical. And surely, not all that is practical in science is beautiful. Here we have both. With all its seemingly abstruse mathematics, the continuous-time model has nevertheless found its way into the mainstream of finance practice. Perhaps its most visible influence on practice has been in the pricing and hedging of financial instruments, an area that has experienced an explosion of real-world innovations over the last decade. In fact, much of the applied research on using the continuous-time model in this area now takes place within practicing financial institutions."

Robert Merton, "Continuous-Time Finance," Blackwell Publishers, (1990).

Portfolio optimization is said to be "myopic" when the investor does not know what will happen beyond the immediate next period. In this framework, basic results about one-period portfolio optimization, such as mean-variance analysis, were described in Part II. Such an approach can be justified for short-term horizons without portfolio rebalancing, or for special utility functions such as power utilities (portfolio choice is myopic when the relative risk aversion is constant and returns are iid).

However, for long-term investment, the investor can benefit from dynamic portfolio optimization, which allows him to take account of important opportunities:

- The investor can modify the portfolio weighting along the whole management period, contrary to a "buy and hold" strategy. Therefore, for example, the portfolio value at maturity can be a quite general function of financial indexes (no longer necessarily linear).

- The investor can use the information delivered at any time by observing financial or economic indicators. In particular, the portfolio strategy can take into account variations of main financial indices.

Despite the unrealistic assumption that a portfolio is actually rebalanced in continuous-time, this approximation can be justified by several reasons:

- First, looking at rebalancing times during a time period $[0, T]$ (for example, intraday market), we observe that they do not correspond to the same deterministic moments, but look like marked point process with values in the whole time interval $[0, T]$.

- Second, when examining financial market properties, we can fix a time scaling so that derivatives hedging and pricing seem to be in continuous-time, and perhaps choose another time scaling in order to assume that portfolio strategies are in continuous-time.

- Finally, under some mild assumptions, discrete-time financial models converge to continuous-time ones, in particular, optimal portfolio strategies (see, *e.g.*, Prigent [414]).

  Indeed, continuous-time modelling leads to more complexity than one-period modelling, requiring introduction of stochastic processes, Ito lemma, martingales, *etc.* However, for standard models, dynamic completeness leads to significant simplification. For utility maximization, explicit solutions can be deduced and analyzed, while, for an one-period incomplete financial market, this analysis may be more involved.

This part is devoted to continuous-time optimization:

- Chapter 6 provides a "brief" summary of standard dynamic optimization. The two main approaches are illustrated by basic examples:

  - The first one is based on the dynamic programming method, using the Pontryagin and Bellman principles.

  - The second one is based on martingale methods and duality.

- Chapter 7 is devoted to the search of optimal payoff profiles and long-term portfolio management:

  - Search of an optimal portfolio value, which is assumed to be a function of a given security price, for example a financial index.

  - Determination of a long-term portfolio, which is composed of three main assets: cash, bond, and stock.

- Finally, Chapter 8 describes main results of financial portfolio optimization when "frictions" are considered:

  - Market incompleteness and/or convex constraints;

  - Transaction costs; and,

  - Other extensions, such as the existence of a labor income or a random time horizon.

Chapter 6

Dynamic programming optimization

6.1 Control theory

In what follows, we consider calculation of variations in the deterministic framework to introduce the main ideas of dynamic programming.

6.1.1 Calculus of variations

6.1.1.1 General problem of the calculus of variations

Very often, this problem takes the following form:

- Let $[0, T]$ be the time period.
- Let U be a function w.r.t. three variables (t, x, y) with real values, assumed to be continuously-differentiable on $[0, T] \times \mathbb{R}^d \times \mathbb{R}^d$.
- We search for a function $x(.)$, continuously-differentiable on $[0, T] \times \mathbb{R}^d$ solution of the optimization problem $(\mathcal{P})$:

$$\max \, \mathcal{U}(x(.)) = \int_0^T U(t, x(t), \dot{x}(t))dt,$$

$$\text{under} \ : \ x(0) = x_0 \text{ and } x(T) = x_T, \tag{6.1}$$

where $\dot{x}(t)$ denotes the derivative of $x(.)$ w.r.t. the current time, and x_0 and x_T are given.

In what follows, the general results are illustrated by some basic examples.

6.1.1.2 Standard consumption-saving problems

Example 6.1
The function $x(.)$ denotes the wealth invested on a riskless asset with rate of return r. Suppose that there exists an income $i(.)$ which is divided into savings $s(.)$ and consumption $c(.)$: $i(t) = s(t) + c(t)$. Then, the wealth dynamics are given by:

$$\dot{x}(t) = rx(t) + s(t). \tag{6.2}$$

Assume that, at a given initial date, the wealth $x(0)$ is equal to x_0, and that the goal at horizon T is to hold x_T. Suppose that $U(t,x,y) = e^{-\rho t}u(x)$, where ρ denotes a psychological discount factor, and $u(.)$ the instantaneous utility (*i.e.*, on consumption). Then, the optimization problem is:

$$\max \int_0^T e^{-\rho t} u(t, x(t))dt,$$
$$\text{under}: \; x(0) = x_0 \text{ and } x(T) = x_T. \tag{6.3}$$

Since $c(t) = i(t) + rx(t) - \dot{x}(t)$, then $U(t,x,y) = e^{-\rho t}u(i(t) + rx - y)$. $\square$

6.1.1.2.1 Euler equation We search for a solution $x(.)$ of the previous problem $(\mathcal{P})$ (with the same assumptions).

PROPOSITION 6.1

(Euler equation)

If $x^*(.)$ is the solution of $(\mathcal{P})$, then it satisfies the equation:

$$\frac{d}{dt}\left(\frac{\partial U}{\partial y}\left(t, x^*(t), \dot{x}^*(t)\right)\right) = \frac{\partial U}{\partial x}\left(t, x^*(t), \dot{x}^*(t)\right). \tag{6.4}$$

PROOF Let $x^*(.)$ be the solution of $(\mathcal{P})$, and let $x(.)$ be another function such that $x(0) = x(T) = 0$. Then, for any $s > 0$, the function $t \to (x^*(t) + sx(t))$ satisfies the constraints of $(\mathcal{P})$. Therefore, we have:

$$\frac{1}{s}\left[\mathcal{U}(x^* + sx) - \mathcal{U}(x^*)\right] \leq 0.$$

Taking the limit when $s \to 0^+$, we deduce:

$$\int_0^T \left[\alpha(t)x(t) + \beta(t)\dot{x}(t)\right] dt \leq 0,$$

with

$$\alpha(t) = \frac{\partial U}{\partial x}\left(t, x^*(t), \dot{x}^*(t)\right); \; \beta(t) = \frac{\partial U}{\partial y}\left(t, x^*(t), \dot{x}^*(t)\right).$$

Integrating by part and using $x(0) = x(T) = 0$, there exists a scalar c such that

$$\int_0^T \left[\beta(t) - \left(c + \int_0^t \alpha(u)\right) du\right] \dot{x}(t)dt \leq 0.$$

Now, consider the particular function $x(t) = \int_0^t \left[\beta(v) - \int_0^v \alpha(u)du\right] dv$ with c such that $\int_0^T \left[\beta(v) - \left(c + \int_0^v \alpha(u)\right) du\right] dv = 0$.

Then, we have:

$$x(0) = x(T) = 0;\ \beta(t) - \left(c + \int_0^t \alpha(u)\right) = \overset{\cdot}{x}(t) \text{ and } \int_0^T \left[\overset{\cdot}{x}(t)\right]^2 dt \leq 0.$$

Thus, necessarily $\overset{\cdot}{x}(t) = 0$ and $\overset{\cdot}{\beta}(t) - \alpha(t) = 0$, which is the Euler equation. ∎

6.1.1.2.2 Application to the consumption-saving plan (see Demange and Rochet [156])

Since $U(t,x,y) = e^{-\rho t} u\left[i(t) + rx - y\right]$ and $c^*(t) = i(t) + rx^*(t) - \overset{\cdot}{x}^*(t)$, we get:

$$\begin{aligned}
\frac{\partial U}{\partial x}\left(t, x^*(t), \overset{\cdot}{x}^*(t)\right) &= re^{-\rho t} u'\left[c^*(t)\right], \\
\frac{\partial U}{\partial y}\left(t, x^*(t), \overset{\cdot}{x}^*(t)\right) &= -e^{-\rho t} u'\left[c^*(t)\right].
\end{aligned}$$

Thus, the Euler equation is:

$$\frac{d}{dt}\left(-e^{-\rho t} u'\left[c^*(t)\right]\right) = re^{-\rho t} u'\left[c^*(t)\right].$$

Then:

$$\begin{aligned}
-\rho u'\left[c^*(t)\right] + \frac{d}{dt}\left(u'\left[c^*(t)\right]\right) = ru'\left[c^*(t)\right] &\Longleftrightarrow \\
\frac{d}{dt}\left(\ln u'\left[c^*(t)\right]\right) = \rho - r &\Longleftrightarrow \\
u'\left[c^*(t)\right] = k \exp\left[(\rho - r)\, t\right]&,
\end{aligned}$$

where k is a non-negative constant. Assuming that u' has an inverse function $j(.)$, we have:

$$c^*(t) = j\left[k \exp\left[(\rho - r)\, t\right]\right]. \tag{6.5}$$

Therefore, the optimal consumption is a monotonous function of current time. It is increasing if and only if the interest rate r is higher than the psychological discount rate ρ.

When the solutions are not necessarily continuously differentiable but "sufficiently" regular, a modified version of the Euler equation can be deduced. Examine solutions which are continuously differentiable, except at a countable number of points where they have left-hand and right-hand derivatives. Denote this set by E.

PROPOSITION 6.2

(Piece-wise differentiable solutions) If $x^*(.)$ belongs to E, then we have:

$$\frac{\partial U}{\partial y}\left(t, x^*(t), \dot{x^*}(t)\right) = \int_0^T \frac{\partial U}{\partial x}\left(t, x^*(t), \dot{x^*}(t)\right) dt + \text{ constant}, \tag{6.6}$$

with:

- If $x^*(.)$ is differentiable at t, we recover the Euler equation.

- If $x^*(.)$ is not differentiable at t, $\frac{\partial U}{\partial y}\left(t, x^*(t), \dot{x^*}(t)\right)$ is continuous (Erdman-Weierstrass condition)

If the terminal value x_T is relaxed, the optimal solution $x^*(.)$ obviously corresponds to a solution of the previous problem $(\mathcal{P})$ with $x_T = x^*(T)$. Therefore, this solution must satisfy the Euler and Erdman-Weierstrass conditions. Nevertheless, since this terminal value is not fixed, it must be compared with all solutions of $(\mathcal{P})$ when x_T is varying. We can also introduce an additionnal function $A(.)$ defined on the terminal value and continuously differentiable. This leads to the following condition:

PROPOSITION 6.3

(Transversality condition)

If $x^{**}(.)$ is solution of the following problem $(\mathcal{P}')$:

$$\begin{aligned} \max\ \mathcal{U}(x(.)) &= \int_0^T U(t, x(t), \dot{x}(t))dt + A\left[x(T)\right], \\ \text{under } &: x(0) = x_0, \end{aligned} \tag{6.7}$$

then

$$\frac{\partial U}{\partial y}\left(t, x^{**}(t), \dot{x^{**}}(t)\right) = -A'\left[x^{**}(t)\right] - \int_0^T \frac{\partial U}{\partial x}\left(t, x^{**}(t), \dot{x^{**}}(t)\right) dt, \tag{6.8}$$

with

- If $x^{**}(.)$ is differentiable at t, we recover the Euler equation:

$$\frac{d}{dt}\left(\frac{\partial U}{\partial y}\left(t, x^{**}(t), \dot{x^{**}}(t)\right)\right) = \frac{\partial U}{\partial x}\left(t, x^{**}(t), \dot{x^{**}}(t)\right). \tag{6.9}$$

- If $x^{**}(.)$ is not differentiable at t, $\frac{\partial U}{\partial y}\left(t, x^{**}(t), \dot{x^{**}}(t)\right)$ is continuous (Erdman-Weierstrass condition).

- At terminal value:

$$\frac{\partial U}{\partial y}\left(t, x^{**}(t), \dot{x^{**}}(t)\right) + -A'\left[x^{**}(t)\right] = 0.$$

PROOF The proof is similar to the previous one: let $x^{**}(.)$ be a solution of problem $(\mathcal{P}')$, and $x(.)$ be a function (continuously differentiable except at a countable number of points) such that $x(0) = 0$. Then, for any $s > 0$, we have:

$$\frac{1}{s}\left[\mathcal{U}\left(x^{*\mu} + sx\right) - \mathcal{U}\left(x^{*\mu}\right)\right] \leq 0.$$

∎

PROOF Taking the limit when $s \to 0^+$, we deduce:

$$\int_0^T \left[\alpha(t)x(t) + \beta(t)\dot{x}(t)\right] dt + \gamma x(T) \leq 0,$$

with

$$\alpha(t) = \frac{\partial U}{\partial x}\left(t, x^*(t), \dot{x}^*(t)\right);\ \beta(t) = \frac{\partial U}{\partial y}\left(t, x^*(t), \dot{x}^*(t)\right);\ \gamma = A'(x(T)).$$

Integrating by part, we have:

$$\int_0^T \left[\beta(t) - \left(\gamma + \int_0^t \alpha(u)\right) du\right] \dot{x}(t)dt \leq 0.$$

Now, consider the particular function $x(t) = \int_0^t \left[\beta(v) - \int_0^v \alpha(u)du\right] dv$, we have: $\int_0^T \left[\dot{x}(t)\right]^2 dt \leq 0$.

Thus, necessarily $\dot{x}(t) = 0$ and $\dot{\beta}(t) = \alpha(t)$, which proves the result.

∎

REMARK 6.1 If the time horizon is no longer fixed, then the transversality condition is modified (see Seierstadt and Sydseter [459]). If the horizon T is infinite, the problem is more involved, since the transversality condition "at infinity" may be not necessary.

∎

***Example 6.2* Consumption-saving problem without fixed wealth at maturity**
(see [156])

$$\max\ \mathcal{U}(x(.)) = \int_0^T e^{-\rho t} u\left(i(t) + rx(t) - \dot{x}(t)\right) dt + A\left[x(T)\right],$$
$$\text{under}\ :\ x(0) = x_0, \tag{6.10}$$

The Euler equation is valid at every point where the control $x(.)$ is continuous. Then:

$$u'(c^{**}(t)) = l \exp\left[(\rho - r)\, t\right].$$

The solution is similar to the previous one when the terminal value is fixed. However, the two constants k and l differ.

Suppose that $u(c) = \ln c$ and $A(x) = \exp(-\rho T) \ln x$. Then,

$$c^*(t) = \frac{1}{k} \exp[(\rho - r)t] \text{ and } c^{**}(t) = \frac{1}{l} \exp[(\rho - r)t].$$

For both cases $c(t) = c^*(t)$ and $c(t) = c^{**}(t)$, the wealth $x(.)$ is given by:

$$\begin{aligned} x(t) &= x_0 e^{(\rho - r)t} + \int_0^t [i(s) - c(s)] e^{r(t-s)} ds, \\ &= x_0 e^{(\rho - r)t} + \int_0^t [i(s)] e^{r(t-s)} ds + \frac{1}{h\rho} e^{rt} \left(1 - e^{-\rho t)}\right), \end{aligned}$$

where $h = k$ or l, according to the optimization problem.

Denote by V_T the discounted total wealth:

$$V_T = x_0 e^{(\rho - r)t} + \int_0^t [i(s)] e^{r(t-s)} ds.$$

- For $(\mathcal{P})$, we get:

$$x_T^* e^{-rT} = V_T + \frac{1}{k\rho} e^{rt} \left(1 - e^{-\rho t}\right).$$

- For $(\mathcal{P}')$, we get:

$$A'(x^{**}(t)) + \frac{\partial U}{\partial y} \left(t, x^{**}(t), \dot{x}^{**}(t)\right) \Longrightarrow \frac{e^{-\rho T}}{x^{**}(T)} - \frac{e^{-\rho T}}{c^{**}(T)} = 0.$$

Thus,

$$\frac{1}{l} e^{-\rho T} = V_T + \frac{1}{l\rho} \left(1 - e^{-\rho T}\right).$$

Finally, the first problem has a rational solution ($k > 0$) if the discounted total wealth V_T is higher than the discounted wealth $x_T^* e^{-rT}$ fixed at time T. Then,

$$\frac{1}{k} = \frac{\left(V_T - X_T e^{-rT}\right) \rho}{\left(1 - e^{-\rho T}\right)}.$$

For the second problem, we deduce:

$$\frac{1}{l} = \frac{V_T \rho}{(\rho - 1) e^{-\rho T} + 1}.$$

$\square$

6.1.2 Pontryagin and Bellman principles

Assume now that the time derivative $\dot{x}(.)$ is no longer a true control variable, but must satisfy an ordinary differential equation (ODE): for t in $[0,T]$,

$$\dot{x}(t) = f\left(t, x(t), v(t)\right), \tag{6.11}$$

where the variable $x(.)$ is the state variable and $v(.)$ is the control variable. Then, the usual optimization problem is:

- Let $[0,T]$ be the time period.
- Let U be a function w.r.t. three variables (t,x,v) with real values, assumed to be continuously differentiable on $[0,T] \times \mathbb{R}^d \times \mathbb{R}^{d'}$.
- Let f be a function w.r.t. three variables (t,x,v) with values in $\mathbb{R}^{d'}$, assumed to be continuously differentiable on $[0,T] \times \mathbb{R}^d \times \mathbb{R}^{d'}$.
- Let A be a function w.r.t. x, with values in $\mathbb{R}$, assumed also to be continuously differentiable on $\mathbb{R}^d$.

We search for a state function $x(.)$ continuously differentiable on $[0,T]$, and a control function $x(.)$, continuous on $[0,T]$ solution of the optimization problem ($\mathcal{P}$"):

$$\begin{aligned} \max\ \mathcal{U}(x(.)) &= \int_0^T U(t, x(t), v(t))dt + A\left[x(T)\right], \qquad (6.12)\\ \text{under}\ :\ \dot{x}(t) &= f\left(t, x(t), v(t)\right),\\ x(0) &= x_0\\ \text{and } v(T) &\in \mathcal{V} \text{ (set of constraints on control)}, \end{aligned}$$

where $\mathcal{V}$ is the set of constraints on the control variable, supposed to be an open convex subset of $\mathbb{R}^{d'}$.

When $f(t,x,v) = v$, the optimization problem is the previous calculus of variations. Otherwise, another method must be introduced. Two "independent" approaches have been proposed:

- *The Pontryagin method* considers that Problem ($\mathcal{P}$") is a calculus of variations with an infinite number of constraints corresponding to the state equation (6.11).
- *The Bellman approach* uses the dynamic programming principle.

However, both use the notion of *Hamiltonian* of $\mathcal{P}$":

DEFINITION 6.1 *The Hamiltonian of* $\mathcal{P}$*" is defined by: for any* (t,x,p,v) *in* $[0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d'}$,

$$H(t,x,p,v) = U(t,x,v) + \langle p, f(t,x,v)\rangle, \tag{6.13}$$

where $\langle ., . \rangle$ *denotes the scalar product on* $\mathbb{R}^d$.

6.1.2.1 The Pontryagin principle

This can be deduced from the Lagrange theorem and the Euler equation, as shown in what follows. For this purpose, define the Lagrange multiplier $p(t)$ associated to the constraint (6.11) at current time t. The Lagrangian $\mathcal{L}$ of $\mathcal{P}$" is:

$$\mathcal{L} = \int_0^T \left[U(t, x(t), v(t)) + \langle p(t), f(t, x(t), v(t)) \rangle - \left\langle p(t), \dot{x}(t) \right\rangle \right] dt + A\left[x(T)\right], \tag{6.14}$$

which is equivalent to:

$$\mathcal{L} = \int_0^T \left[H(t, x(t), p(t), v(t)) - \left\langle p, \dot{x}(t) \right\rangle \right] dt + A\left[x(T)\right]. \tag{6.15}$$

Therefore, the maximization w.r.t. $(x, \dot{x}, v)$ can be decomposed into two steps:

- Step 1: Maximization w.r.t. v.

  Any optimal control must satisfy the following condition:

  $$v^*(t) = \arg\max_v H(t, x(t), p(t), v(t)), \ \forall t \text{ a.s.}$$

  Denote $H^*(t, x(t), p(t), v(t))$ the maximum of the Hamiltonian.

- Step 2: Maximization w.r.t. $(x, \dot{x})$.

  We have to solve the following problem of calculus of variations: for $x(0) = x_0$,

  $$\max \int_0^T \left[H^*(t, x(t), p(t), v(t)) - \left\langle p(t), \dot{x}(t) \right\rangle \right] dt + A\left[x(T)\right].$$

  Since the terminal value x_T is not fixed, the optimality conditions are given by Proposition 6.4:

- If $v(.)$ is continuous at t, we recover the Euler equation:

$$\dot{p}(t) = -\frac{\partial H^*}{\partial x}(t, x(t), p(t)). \tag{6.16}$$

- If $v(.)$ is discontinuous at t, $t \to p(t)$ is continuous (Erdman-Weierstrass condition).

- At terminal value: (transversality condition)

$$p(T) = A'(x(T)).$$

$$\dot{x}(T) = -\frac{\partial H^*}{\partial p}(t, x(t), p(t)).$$

Therefore, we have to solve the system of Hamilton-Jacobi equations:

$$\dot{x}(T) = -\frac{\partial H^*}{\partial p}(t, x(t), p(t)), \tag{6.17}$$

$$\dot{p}(T) = -\frac{\partial H^*}{\partial x}(t, x(t), p(t)), \tag{6.18}$$

with $p(T) = A'(x(t))$ and $x(0) = x_0$.

Note that:
- When H^* does not depend on current time, the Hamiltonian is constant for solutions of:

$$\frac{d}{dt}\left[H^*(x(t), p(t))\right] = \frac{\partial H^*}{\partial x}\dot{x} + \frac{\partial H^*}{\partial p}\dot{p} = 0.$$

- The boundary condition implies that both the initial values of $x(.)$ and the terminal value of $p(.)$ are fixed.

THEOREM 6.1 **Pontryagin principle**

Assume that:
- The utility function U, the function f, and the function A are continuously differentiable on their respective domains.
- In addition, the function f is bounded and Lipschitzian w.r.t. x, uniformly w.r.t. (t, v).
- The set $\mathcal{V}$ is a convex and compact subset of $\mathbb{R}^{d\prime}$.
Then, if $(x^*(.), v^*(.))$ is solution of $\mathcal{P}$", there exists $p^*(.)$, belonging to E, such that:
1) $v^*(t) = \arg\max_v H(t, x^*(t), p^*(t), v)$, $\forall t$ a.s. Denote this maximum by $H^*(t, x^*(t), p^*(t), v)$.
2) The pair (x^*, p^*) is the solution of the Hamilton-Jacobi equation:

$$\begin{cases} \dot{x}(T) = -\frac{\partial H^*}{\partial p}(t, x(t), p(t)), \\ \dot{p}(T) = -\frac{\partial H^*}{\partial x}(t, x(t), p(t)), \end{cases} \textit{with} \begin{cases} p(T) = A'(x(t)), \\ x(0) = x_0. \end{cases} \tag{6.19}$$

REMARK 6.2 This theorem is applied through two steps:
1) Hamiltonian maximization: we search for $v^*(t, x(t), p(t))$, which determines $H^*(t, x(t), p(t), v)$.
2) Hamilton-Jacobi equations solution: we search for (x^*, p^*), which determines $v^*(t, x^*(t), p^*(t))$.

$\square$

Example 6.3 **Deterministic portfolio optimization with transaction costs**
(see [156])

Assume that an individual invests his wealth on cash with a rate of return $r(t)$, and stock with price $S(t)$ and dividend rate $d(t)$.

Each transaction (buying or selling) has a proportional cost θ and there exists an upper bound N on the number of traded stocks.

Denote: $C(t)$ the cash amount, $n(t)$ the number of purchased stocks, and $q(t)$ the number of sold stocks.

The state variable is $x(t) = (C(t), n(t))$ and the control variable is $v(t) = q(t)$.

The optimization problem is:

$$\begin{aligned} &\max\left(C(T) + n(T)S(T)\right) \\ \dot{C} &= rV + dn + S(q - \theta\,|q|), \\ \dot{n} &= -q, \\ C(0) &= C_0 \text{ and } n(0) = p_0, \\ q &\in [-N, N]. \end{aligned}$$

The Hamiltonian does not depend on current time and is given by:

$$H(x, p, v) = H(C, n, p_1, p_2, q) = p_1\left(rV + dn + S(q - \theta\,|q|)\right) - p_2 q.$$

Step 1: Hamiltonian maximization, w.r.t. q on $[-N, N]$.
The solution is given by:

$$\begin{cases} q^* = N \text{ if } \ p_1 S(1-\theta) > p_2, \\ q^* = 0 \text{ if } p_1 S(1-\theta) < p_2 < \ p_1 S(1+\theta), \\ q^* = -N \text{ if } p_2 > p_1 S(1+\theta), \end{cases}$$

and

$$H^*(x, p) = p_1\left(rV + dn\right) + N\max\left(p_1 S(1-\theta) - p_2, p_2 - p_1 S(1+\theta), 0\right).$$

Step 2: Hamilton-Jacobi equations solution.

$$\begin{cases} \dot{p}_1 = -\frac{\partial H^*}{\partial V} = -rp_1, \\ \dot{p}_2 = -\frac{\partial H^*}{\partial n} = -dp_1, \end{cases} \text{with } \begin{cases} p_1(T) = 1, \\ p_2(T) = S(T). \end{cases}$$

The solution is given by:

$$\begin{cases} p_1 = \exp\left[\int_t^T r(s)ds\right], \\ p_2 = S(T) + \int_t^T d(s)p_1(s)ds. \end{cases}$$

Therefore, $p_1(t)$ can be viewed as the value at time T of one monetary unit invested on time t, invested on the cash. $p_2(t)$ is the value at time T of one stock invested from t to T.

The optimal control is "bang-bang" : it can have three values $-N, 0$, or N and is piece-wise constant. Its value depends on the ratio $\tau = p_2/p_1 S$, which is the cuurent value of one monetary unit invested on stock from t to T. We have:

$$\tau(t) = \frac{\exp\left[-\int_t^T r(s)ds\right]\left(S(T) + \int_t^T d(s)p_1(s)ds\right)}{S(t)}.$$

Thus, the optimal control satisfies:

$$\begin{cases} q^* = N \text{ if } \tau(t) < (1-\theta), \\ q^* = 0 \text{ if } (1-\theta) < \tau(t) < (1+\theta), \\ q^* = -N \text{ if } \tau(t) > (1+)\,\theta. \end{cases}$$

Note that if there exists no friction ($\theta = 0$ and $N = \infty$), then $\tau(t)$ must be equal to 1 in order to have a solution. In that case, we recover the standard no-arbitrage valuation:

$$S(t) = \exp\left[-\int_t^T r(s)ds\right]\left(S(T) + \int_t^T d(s)p_1(s)ds\right).$$

□

6.1.2.2 The Bellman principle

The Bellman approach considers that problem ($\mathcal{P}$") can be embedded in a more general class, parametrized by the initial date t_0 and the initial state value x_{t_0}, solved by the dynamic programming principle.

Consider the functional $\mathcal{J}$ defined by:

$$\mathcal{J}(t_0, x_{t_0}) = \max_{v(.)} \int_{t_0}^T U(t, x(t), v(t))dt + A\left[x(T)\right], \tag{6.20}$$

$$\begin{aligned} &\text{under :} \\ &\dot{x}(t) = f\left(t, x(t), v(t)\right) \text{ if } v(.) \text{ is continuous at } t, \qquad (6.21) \\ &x(t_0) = x_0 \text{ and } v(T) \in \mathcal{V}\ , \end{aligned}$$

where $\mathcal{V}$ is the set of piece-wise continuous functions.

- Assume that the ODE $\dot{x}(t) = f\left(t, x(t), v(t)\right)$ has one and only one solution.

- Suppose that (t_0, x_{t_0}) belongs to the domain $\mathcal{D}$ where $\mathcal{J}$ is defined. Consider the subdivision of the interval $[t_0, T]$ into $[t_0, t_0 + h[$ and $[t_0 + h, T]$.

The optimization problem can be divided into two steps:

- Maximization on $[t_0+h, T]$, knowing the state value at time t_0+h. Denote the optimal value by $\mathcal{J}(t_0+h, x_{t_0+h})$.

- Maximization on $[t_0, t_0+h[$, taking account of its impact on $\mathcal{J}(t_0+h, x_{t_0+h})$.

This leads to the dynamic programming principle:

$$\mathcal{J}(t_0, x_{t_0}) = \max_{v(.)} \int_{t_0}^{t_0+h} U(t, x(t), v(t))dt + \mathcal{J}(t_0+h, x_{t_0+h}). \tag{6.22}$$

If (t_0, x_{t_0}) belongs to the interior of the domain $\mathcal{D}$, $\mathcal{J}(t_0, x_{t_0})$ exists for sufficient small values of h and we have:

$$0 = \max_{v(.)} \frac{1}{h}\left[\int_{t_0}^{t_0+h} U(t, x(t), v(t))dt + \mathcal{J}(t_0+h, x_{t_0+h}) - \mathcal{J}(t_0, x_{t_0})\right]. \tag{6.23}$$

Assuming that $\mathcal{J}$ is differentiable at (t_0, x_{t_0}), we deduce (for $h \to 0^+$):

$$0 \geq \max_{v(.)}\left[U(t_0, x_{t_0}, v) + \frac{\partial}{\partial t}\mathcal{J}(t_0, x_{t_0}) + \frac{\partial}{\partial x}\mathcal{J}(t_0, x_{t_0}) f\,(t_0, x_{t_0}, v)\right],$$

$$0 \geq \frac{\partial}{\partial t}\mathcal{J}(t_0, x_{t_0}) + \max_{v(.)}\left[U(t_0, x_{t_0}, v) + \left\langle \frac{\partial}{\partial x}\mathcal{J}(t_0, x_{t_0}), f\,(t_0, x_{t_0}, v)\right\rangle\right].$$

Therefore, using the Hamiltonian, we have:

$$0 \geq \frac{\partial}{\partial t}\mathcal{J}(t_0, x_{t_0}) + H^*(t_0, x_{t_0}, \frac{\partial}{\partial x}\mathcal{J}(t_0, x_{t_0})). \tag{6.24}$$

Under mild assumptions, the reverse inequality can be proved. Thus, for varying t and x, $\mathcal{J}$ satisfies the Bellman partial derivative equation (PDE).

THEOREM 6.2 Bellman PDE

If the value function $\mathcal{J}$ associated to problem ($\mathcal{P}$") is defined and continuously differentiable on $]t_0, T[\times\mathcal{O}$ where $\mathcal{O}$ is a convex open subset of $\mathbb{R}^d$, then it satisfies the Bellman PDE:
For any $(t, x) \in]t_0, T[\times\mathcal{O}$,

$$\frac{\partial}{\partial t}\mathcal{J}(t, x) + H^*(t, x, \frac{\partial}{\partial x}\mathcal{J}(t, x)) = 0 \text{ with } \mathcal{J}(T, x) = A(x). \tag{6.25}$$

REMARK 6.3 The solution is deduced through two steps:

- Hamiltonian maximization.

- Bellman PDE solution.

The Bellman approach allows recovery of the maximum principle and the Hamilton-Jacobi equations, introduced in the Pontryagin approach.

Indeed, consider $p(t) = \frac{\partial}{\partial x}\mathcal{J}(t, x(t))$ and assume that $\mathcal{J}$ is twice-continuously differentiable. Then, we have to check that $p(.)$ is solution of the Hamilton-Jacobi ODE, and also that $p(.)$ satisfies the transversality condition.

1) $\underline{p(.)\text{ is solution of the Hamilton-Jacobi ODE:}}$

Since $p(t) = \frac{\partial}{\partial x}\mathcal{J}(t, x(t))$, by differentiating $p_i(t)$, we get:

$$\dot{p_i}(t) = \frac{d}{dt}\left[\frac{\partial \mathcal{J}}{\partial x_i}(t, x(t))\right] = \frac{\partial^2 \mathcal{J}}{\partial t \partial x_i}(t, x(t)) + \sum_{j=1}^{d} \frac{\partial^2 \mathcal{J}}{\partial x_i \partial x_j}(t, x(t)) .\dot{x_j}(t).$$

Differentiating the Bellman equation $\frac{\partial \mathcal{J}}{\partial t} + H^*(t, x, \frac{\partial \mathcal{J}}{\partial x}) = 0$ w.r.t. x_i, we deduce:

$$\frac{\partial^2 \mathcal{J}}{\partial t \partial x_i} + \frac{\partial H^*}{\partial x_i} + \sum_{j=1}^{d} \frac{\partial^2 \mathcal{J}}{\partial x_i \partial x_j}.\frac{\partial H^*}{\partial x_i} = 0.$$

Finally, since $f_j(t, x, v^*) = \frac{\partial H^*}{\partial p_j}\left(t, x, \frac{\partial \mathcal{J}}{\partial x}\right)$ and $\dot{x_i} = f_j(t, x, v^*)$, the result is proved:

$$\dot{p_i}(t) = -\frac{\partial H^*}{\partial x_i}(t, x(t), p(t)).$$

2) $\underline{p(T)\text{ satisfies the transversality condition:}}$

Differentiating w.r.t. x_i, the terminal value of Bellman equation, we have:

$$\frac{\partial \mathcal{J}}{\partial x_i}(T, x) = \frac{\partial A}{\partial x_i}(x),$$

which implies:

$$p(T) = \frac{\partial \mathcal{J}}{\partial x}(T, x(T)) = \frac{\partial A}{\partial x}(x(T)).$$

∎

REMARK 6.4 The Bellman method allows us to more easily find the solution when its form is anticipated. Otherwise, computation difficulties of the Bellman PDE and Hamilton-Jacobi ODE are often similar.

∎

6.1.3 Stochastic optimal control

6.1.3.1 Introduction to stochastic optimal control

The theory of stochastic optimal control extends previous methods to the stochastic dynamical systems. For basic notations, definitions and properties of stochastic processes, we refer to Karatzas and Shreve ([320],[321]) and Jacod and Shiryaev [294]. Appendix B provides a short survey about these notions.

Consider the case of systems associated to stochastic differential equations (SDE):

$$dX_t = f(t, X_t)dt + g(t, X_t)dW_t \text{ with } X_{t_0} = x_0, \tag{6.26}$$

where X_t is the state variable with values in $\mathbb{R}^d$, $v(t)$ is the control variable with values in $\mathbb{R}^{d'}$, and W is a d"$-$dimensional Brownian motion with d" $\leq d$.

The functions $f(.,.)$ and $g(.,.)$ are assumed to satisfy usual conditions in order to ensure that the previous SDE has one and only one solution. For example:
- The functions $f(.,.)$ and $g(.,.)$ have linear growth:

$$||f(t,x)|| \leq (1 + ||x||)\, M_t \text{ and } ||g(t,x)|| \leq (1 + ||x||)\, M_t,$$

where $M(.)$ is a positive determistic function, upper bounded on each compact subset of $[0, T]$.

- The functions $f(.,.)$ and $g(.,.)$ are Lipschitzian:

$$||f(t,x) - f(t,y)|| \leq (||x - y||)\, K_t \text{ and } ||g(t,x) - g(t,y)|| \leq (||x - y||)\, K_t,$$

where $K(.)$ is also a positive determistic function, upper bounded on each compact subset of $[0, T]$.

Due to the observation of the path on time period $[0, T]$, the acquired information may modify the control variable $v(.)$. Thus, we assume that $v(.)$ is now a stochastic process $(v(t, \omega))_t$. Furthermore, if $v(.)$ is supposed to be such that $v(t, \omega) = \widetilde{v}(t, X_t(\omega))$, where $\widetilde{v}$ is a deterministic function $\widetilde{v} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$, then $v(.)$ is said to be "feedback" and the system is Markovian:

$$dX_t = f(t, \widetilde{v}(t, X_t))dt + g(t, \widetilde{v}(t, X_t))dW_t \text{ with } X_{t_0} = x_0. \tag{6.27}$$

The information is modelled by the filtration $\mathcal{F}_t^X$ generated by process X.

The matrix $g(t, x)$ is assumed to satisfy the following: There exists $\varepsilon > 0$ such that $Trace\left({}^t g(t, x) g(t, x)\right) \geq \varepsilon$ (non-degenerated).

Under previous assumptions, we are led to the following stochastic control problem $\mathcal{P}_c$:

$$\max_{v(.)} \mathbb{E}\left[\int_{t_0}^{T} U(t, X(t), v(t))dt + A\left[X(T)\right]|X(t_0) = x_0\right], \quad (6.28)$$

under, for $t_0 \leq t \leq T$,

$$\begin{aligned} dX_t &= f(t, v(t))dt + g(t, v(t)dW_t, \quad (6.29)\\ X(t_0) &= x_0 \text{ and } v(t) = \widetilde{v}(t, X_t(\omega)) \in \mathcal{V} \ a.s. \ , \end{aligned}$$

where $\mathcal{V}$ is an open convex subset of $\mathbb{R}^{d'}$.

The objective function U is assumed to be twice-differentiable on $[t_0, T] \times \mathbb{R}^d \times \mathbb{R}^{d'}$, and A is twice-differentiable on $\mathbb{R}^d$.

The maximization is made on functions $\widetilde{v}$ which are piece-wise continuous w.r.t. the current time t and Lipschitzian w.r.t. the state variable x.

Problem ($\mathcal{P}_c$) is embedded in the class parametrized by (t, x). Thus, we can apply the Bellman approach. For this purpose, we define the new Hamiltonian.

DEFINITION 6.2 *The Hamiltonian $H(t, x, p, q, v)$ associated to problem ($\mathcal{P}_c$) is given by: for $(t, x, p, q, v) \in [t_0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times (\mathbb{R}^d \times \mathbb{R}^d) \times \mathbb{R}^{d'}$,*

$$H(t, x, p, q, v) = U(t, x, v) + \sum_{i=1}^{d} p_i f_i(x, v) + \frac{1}{2} \sum_{i,l=1}^{d} q_{i,l} \left(\sum_{j=1}^{d"} \sigma_{i,j}(x, v)\sigma_{l,j}(x, v) \right). \quad (6.30)$$

With respect to the deterministic case, any pair of state variables (x_i, x_l) is associated a new variable $q_{i,l}$ with a coefficient equal to half the instantaneous covariance of the random variables X_i and X_l. This term is due to the Dynkin operator of the process X, denoted by $\mathcal{D}(X)$. Note that $\mathcal{D}(X)$ can be viewed as a "mean or expectation of variation rate" of the process X, "roughly" equal to the ordinary derivative w.r.t. current time as for deterministic functions:

$$\frac{\mathbb{E}[dX_t \,|\mathcal{F}_t]}{dt}.$$

The Hamiltonian can also be expressed as:

$$U(t, x, v) + \langle p, f(x, v)\rangle + \frac{1}{2} Trace\left(g(t, x)q^t g(t, x)\right).$$

6.1.3.2 The stochastic Bellman principle

We apply the dynamic programming approach to the stochastic framework.

Denote $\mathcal{J}(t_0, x_0)$ as the value function associated to problem $\mathcal{P}_c$.

Assume that (t_0, x_0) belongs to the interior domain of the operator $\mathcal{J}$.

This leads to the stochastic dynamic programming principle:

$$\mathcal{J}(t_0, x_{t_0}) = \max_{\widetilde{v}(.)} \mathbb{E}\left[\int_{t_0}^{t_0+h} U(t, X(t), v(t))dt + \mathcal{J}(t_0+h, X_{t_0+h})\,|X_{t_0} = x_0\right]. \tag{6.31}$$

Then:

$$0 = \max_{v(.)} \left\{ \begin{array}{c} \frac{1}{h}\mathbb{E}\left[\int_{t_0}^{t_0+h} U(t, x(t), v(t))dt\,|X_{t_0} = x_0\right] + \\ +\frac{1}{h}\mathbb{E}\left[\mathcal{J}(t_0+h, X_{t_0+h}) - \mathcal{J}(t_0, X_{t_0})\,|X_{t_0} = x_0\right] \end{array} \right\}, \tag{6.32}$$

where the optimum is searched on the set of feedback controls $\widetilde{v}$, and the process X is solution of the SDE:

$$dX_t = f(t, \widetilde{v}(t, X_t))dt + g(t, \widetilde{v}(t, X_t)dW_t \text{ with } X_{t_0} = x_0. \tag{6.33}$$

For $h \to 0^+$, we deduce:

$$\lim_{h\to 0^+} \frac{1}{h}\mathbb{E}\left[\int_{t_0}^{t_0+h} U(t, x(t), v(t))dt\,|X_{t_0} = x_0\right] = U(t_0, X_{t_0}, v).$$

Then, applying Ito's lemma, we get:

$$\lim_{h\to 0^+} \frac{1}{h}\mathbb{E}\left[\mathcal{J}(t_0+h, X_{t_0+h}) - \mathcal{J}(t_0, X_{t_0})\,|X_{t_0} = x_0\right] = \mathcal{D}\mathcal{J}(t_0, X_{t_0}).$$

Thus, we have:

$$0 \geq \max_{v(.)} \left[U(t_0, x_{t_0}, v) + \frac{\partial \mathcal{J}}{\partial t}(t_0, x_{t_0}) + \left\langle \frac{\partial \mathcal{J}}{\partial x}(t_0, x_{t_0}), f\,(t_0, x_{t_0}, v)\right\rangle\right]$$
$$+\frac{1}{2}Trace\left(g(t_0, x_{t_0}, v)^t g(t_0, x_{t_0}, v)\right).$$

Using mild assumptions, the equality is proved.
Set

$$p = \frac{\partial \mathcal{J}}{\partial x}(t_0, x_{t_0}) \text{ and } q = \frac{\partial^2 \mathcal{J}}{\partial x^2}(t_0, x_{t_0}).$$

Therefore, using the Hamiltonian, we are led to the stochastic Bellman equation.

$$0 = \frac{\partial \mathcal{J}}{\partial t}(t_0, x_{t_0}) + H^*(t_0, x_{t_0}, \frac{\partial \mathcal{J}}{\partial x}(t_0, x_{t_0}), \frac{\partial^2 \mathcal{J}}{\partial x^2}(t_0, x_{t_0})), \tag{6.34}$$

where

$$H^*(t_0, x_{t_0}, \frac{\partial \mathcal{J}}{\partial x}(t_0, x_{t_0}), \frac{\partial^2 \mathcal{J}}{\partial x^2}(t_0, x_{t_0}))$$
$$= max_{v(.)} H(t_0, x_{t_0}, \frac{\partial \mathcal{J}}{\partial x}(t_0, x_{t_0}), \frac{\partial^2 \mathcal{J}}{\partial x^2}(t_0, x_{t_0}), v).$$

Thus, for varying t and x, $\mathcal{J}$ satisfies the stochastic Bellman equation.

THEOREM 6.3 Stochastic Bellman equation

If the value function $\mathcal{J}$ associated to problem ($\mathcal{P}_c$) is defined and continuously differentiable on $]t_0, T[\times\mathcal{O}$ where $\mathcal{O}$ is a convex open subset of $\mathbb{R}^d$, then it satisfies the Bellman equation: for any $(t, x) \in]t_0, T[\times\mathcal{O}$,

$$\frac{\partial \mathcal{J}}{\partial t}(t, x) + H^*(t, x, \frac{\partial \mathcal{J}}{\partial x}(t, x), \frac{\partial^2 \mathcal{J}}{\partial x^2}(t, x)) = 0 \text{ with } \mathcal{J}(T, x) = A(x). \quad (6.35)$$

REMARK 6.5 The solution is again deduced through two steps:

- Hamiltonian maximization:

$$H^*(t, x, p, q) = \max_{v(.)} H(t, x, p, q, v), and$$

- Solution of the Bellman equation:

For a solution $v^*(.)$ of previous optimization problem, we have to solve the Bellman PDE:

$$\frac{\partial \mathcal{J}}{\partial t}(t, x) + H^*(t, x, \frac{\partial \mathcal{J}}{\partial x}(t, x), \frac{\partial^2 \mathcal{J}}{\partial x^2}(t, x)) = 0 \text{ with } \mathcal{J}(T, x) = A(x).$$

Under previous assumptions, there exists a maximal domain on which there exists one and only one solution. However, solutions are rarely explicit.

□

For stationary problems, explicit solutions may be determined.

Consider, for example, the case of an infinite horizon and suppose that the investor has an exponential utility function (CARA utility).

Example 6.4
Suppose that $T = \infty$ and U is such that:

$$U(t, x, v) = e^{-\rho t} u(x, v),$$

with $\rho > 0$.

Suppose also that $d = 1$ (univariate case).

Consider

$$\mathcal{J}(t, x) = \max_{\widetilde{v}(.)} \mathbb{E}\left[\int_t^{\infty} e^{-\rho s} u(X(s), \widetilde{v}(s)) ds \,| X_t = x\right], \tag{6.36}$$

with

$$dX_s = f(X_s, \widetilde{v}(s, X_s))ds + g(X_s, \widetilde{v}(s, X_s))dW_s, \tag{6.37}$$

$$X(t) = x \text{ and } v(s) = \widetilde{v}(s, X_s) \in \mathcal{V} \ .$$

Then, if $\mathcal{J}(0, x) = I(x)$ is defined and twice-continuously differentiable on the open subset $\mathcal{O}$, $\mathcal{J}(t, x)$ is defined on the whole space $\mathbb{R} \times \mathcal{O}$ and we have:

For any $(t, x) \in \mathbb{R} \times \mathcal{O}$,

$$\mathcal{J}(t, x) = e^{-\rho t} I(x),$$

where I is solution of the ODE:

$$\rho I(x) = \max_{v \in \mathcal{V}} \left[u(x, v) + I'(x) f(x, v) + \frac{1}{2} I''(x) g^2(x, v)\right].$$

□

The next section illustrates this approach to determine optimal financial portfolios.

6.2 Lifetime portfolio selection

6.2.1 The optimization problem

In what follows, we consider the approach introduced by Samuelson [447], and further studied by Merton ([385], [386] and [387]), who used dynamic programming methods in order to deduce explicit solutions when security parameters are deterministic. A more general solution is also examined when security parameters are no longer deterministic.

Assumptions (S) on securities:

- Let S be the price vector of d securities. Let $(\Omega, \mathbb{P})$ represent the probability space. Assume that it is defined from the following stochastic differential equation (SDE):

$$dS_t = S_{t_-}.\left(\mu(t, S_t)dt + \sigma(t, S_t)dW_t\right), \tag{6.38}$$

where $W = (W_1, ..., W_n)$ is a $d-$multidimensional Brownian motion such that

$$\forall i \neq j,\ Cov(W_{i,t}, W_{j,t}) = \rho_{i,j}t \text{ and } Var(W_{i,t}) = t. \tag{6.39}$$

- The information is modelled by the filtration $\mathcal{F}_t$ generated by the Brownian motion (and, as usual, completed in order to contain all $\mathbb{P}$-null sets).

- The time horizon is supposed to be finite and is denoted by T.

- The processes $\mu(.,.) = (\mu_1, ..., \mu_d)(.,.)$, which model the instantaneous expectations, and $\sigma(.,.) = (\sigma_1, ..., \sigma_d)(.,.)$, which model the volatilities, satisfy usual conditions which ensure that the previous SDE has one and only one solution (see for example [294]). For instance:

i) These processes are measurable, $\mathcal{F}_t$-adapted and uniformly bounded on $[0, T] \times \Omega$.

ii) The matrix $\sigma(t, .)$ is invertible with bounded inverse for any $t \in [0, T]$. The process σ is also predictable.

REMARK 6.6 Under the previous hypothesis, the financial market is complete and without arbitrage opportunity.

□

Assumptions (P) on portfolio strategies:

- At any time t, the investor chooses the positive amount c_t per time unit, which is assigned to his consumption, and also the portfolio weighting w_t.

- The investor's strategy is supposed to be self-financing and the cumulative consumption $\int_0^t c_s ds$ is an $\mathcal{F}_t$-adapted-process with $\int_0^t c_s ds < \infty$, $\mathbb{P}$-a.s.

- The portfolio weighting w_t is predictable and such that $\int_0^t ||w_s||^2 ds < \infty$, $\mathbb{P}$-a.s., where $||.||$ denotes the norm.

Thus, the portfolio value V_t is an Ito process defined by:

$$V_t = V_0 + \sum_{i=1}^{d} \int_0^t w_{i,s} V_s \frac{dS_{i,s}}{S_{i,s}} - \int_0^t c_s ds. \tag{6.40}$$

Assumptions (U) on utility functions:

For any time t, let $U(.,t)$ and $\widetilde{U}(.,t)$ be two utility functions satisfying: $\forall t \in [0,T]$,

- $U(.,t)$ and $\widetilde{U}(.,t)$ are defined on $\mathbb{R}^+$, strictly concave, non-decreasing and continuously differentiable.

- $\lim_{x\to\infty} \frac{\partial}{\partial x} U(x,t) = 0$ and $\lim_{x\to\infty} \frac{\partial}{\partial x} \widetilde{U}(.,t) = 0$.

- $U(.,.)$ and $\widetilde{U}(.,.)$ are continuous on $\mathbb{R}^+ \times [0,T]$ (for example: $U(x,t) = e^{-\rho t} u(x)$ with ρ positive scalar).

Note that, under the first assumption, the marginal utilities $\frac{\partial}{\partial x} U(x,t)$ and $\frac{\partial}{\partial x} \widetilde{U}(.,t)$ are non-decreasing. Therefore, their inverse functions J and $\widetilde{J}$ exist.

The maximization of the expected intertemporal utility along the time period $[0,T]$ is the following optimization problem:

$$\max_{c,w} \mathbb{E}\left[\int_0^T U(c_s,s)ds + \widetilde{U}\left(V_T,T\right)\right]. \tag{6.41}$$

There exists one constraint corresponding to the initial budget: at time $t=0$, the portfolio value V is equal to a given value V_0.

6.2.2 The deterministic coefficients case

The functional to be optimized is time-additive and strategies are assumed to be non-anticipative (*i.e.*, they are functions of past or current information and not of future observations).

Therefore, as seen in the previous section, the solution can be determined through dynamic programming: "an optimal strategy on the whole time period $[0,T]$ must be optimal on any sub-period $[t,T]$. "

Then, at any time t, the consumption rate process C and the weighting process w are solutions of the following problem:

- Consider the functional $\mathcal{J}$ ("the value-function") given by:

$$\mathcal{J}(V,S,t) = \max_{c,w} \ \mathbb{E}_t \left[\int_t^T U(c_s,s)ds + \widetilde{U}\left(V_T,T\right) \right], \tag{6.42}$$

where $\mathbb{E}_t$ denotes the conditional expectation given the information at time t.

- Consider the functional Φ defined by:

$$\Phi(c,w,V,S,t) = U(c_t,t) + \mathcal{D}(\mathcal{J}), \tag{6.43}$$

where $\mathcal{D}$ denotes the Dynkin operator associated to price variable S and with value V for given control variables c and w:

$$\mathcal{D} = \frac{\partial}{\partial t} + \left(\sum_{i=1}^{d} w_{i,t}\mu_{i,t}V_t - c_t\right)\frac{\partial}{\partial V} + \sum_{i=1}^{d}\mu_{i,t}S_{i,t}\frac{\partial}{\partial S_i} \tag{6.44}$$

$$+\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_{i,j,t}w_{i,t}w_{j,t}V_t^2\frac{\partial^2}{\partial V^2} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_{i,j,t}S_{i,t}S_{j,t}V_t^2\frac{\partial^2}{\partial S_{i,t}\partial S_{j,t}}$$
$$+\sum_{i=1}^{d}\sum_{j=1}^{d}S_{i,t}V_t\sigma_{i,j,t}w_{i,t}\frac{\partial^2}{\partial S_{i,t}\partial V_t}.$$

Using standard results of dynamic programming (see Kushner [340]), we deduce:

PROPOSITION 6.4
If U is strictly concave w.r.t. c, and if $\widetilde{U}$ is concave w.r.t V, then there exists a set of optimal controls w^* and c^* such that: $\sum_{i=1}^{d} w_i = 1$, $\mathcal{J}(V,S,T) = \widetilde{U}\left(V_T,T\right)$, and for any t in $[0,T]$,

$$0 = \Phi(c^*,w^*,V,S,t) \geq \Phi(c,w,V,S,t). \tag{6.45}$$

This first result shows that the optimization problem is equivalent to the maximization of the functional $\Phi(c,w,V,S,t)$ under the following constraint

on weights w: $\sum_{i=1}^{d} w_{i,t} = 1$. This can be done by using the Lagrangian:

$$L = \Phi(c, w, V, S, t) + \lambda \left(1 - \sum_{i=1}^{d} w_{i,t}\right),$$

where λ denotes the usual Lagrange parameter. Under previous assumptions, the first-order conditions are given by: For any $i \in \{1, ..., d\}$,

$$0 = L_c(c^*, w^*) = U_c(c^*, t) - \mathcal{J}_V, \tag{6.46}$$

$$0 = L_{w_i}(c^*, w^*) = -\lambda + \mathcal{J}_V \mu_i V + \mathcal{J}_{VV} \sum_{j=1}^{d} \sigma_{i,j} w^*_{i,t} V^2 + \sum_{j=1}^{d} \mathcal{J}_{S_j V} \sigma_{i,j} S_j V,$$

$$0 = 1 - \sum_{i=1}^{d} w_{i,t}. \tag{6.47}$$

(notation: A_X denotes the partial derivative of A w.r.t. X and A_{XY} the partial derivative of order 2 w.r.t. X and Y).
Note that $L_{cc} = U_{cc} < 0$, $L_{cw_i} = 0$, $L_{w_i,w_i} = \sigma_i^2 V^2 \mathcal{J}_{VV}$, $L_{w_i,w_j} = \sigma_{i,j} V^2 \mathcal{J}_{VV}$, and that the volatility matrix $[\sigma_{i,j}]_{i,j}$ is non-degenerate. Therefore, a sufficient condition for the existence of an interior solution is: $\mathcal{J}_{VV} < 0$ (meaning that $\mathcal{J}$ is strictly concave w.r.t. V). By differentiating the first previous equation, we deduce that the optimal consumption is an increasing function w.r.t. V.

Next, we must determinate c^*, w^*, and λ as functions of S, V, and t and the derivatives of $\mathcal{J}$ by solving $(n+2)$ implicit equations. Then, we have to substitute these solutions into Equation (6.45), in order to get a differential equation of the second order w.r.t. $\mathcal{J}$ with boundary condition: $\mathcal{J}(V, S, T) = \widetilde{U}(V_T, T)$.

Denote by J, the inverse of the marginal utility U' w.r.t. the consumption c:

$$J = (U')^{-1}. \tag{6.48}$$

From Equation (6.46), we deduce:

$$c_t^* = J(\mathcal{J}_V, t). \tag{6.49}$$

The optimal weights w_i^* are determined from a linear system, which allows us to get explicit solutions. For this purpose, denote:

$$\Sigma = [\sigma_{i,j}]_{i,j}, \Sigma^{-1} = [\nu_{i,j}]_{i,j} \text{ and } \Gamma = \sum_{i=1}^{d}\sum_{j=1}^{d} \nu_{i,j}.$$

By eliminating the Lagrange parameter λ in the second equation (6.46), we determine the weights w_i^* :

$$w_{i,t}^* = h_i(S_t,t) + m(S_t,V_t,t)g_i(S_t,t) + f_i(S_t,V_t,t), \tag{6.50}$$

with

$$\sum_{i=1}^{d} h_i = 1, \sum_{i=1}^{d} f_i = \sum_{i=1}^{d} g_i = 0,$$

and

$$h_i(S_t,t) = \sum_{j=1}^{d} \frac{\nu_{i,j}}{\Gamma}; m(S_t,V_t,t) = -\frac{\mathcal{J}_V}{V\mathcal{J}_{VV}},$$

$$g_i(S_t,t) = \frac{1}{\Gamma}\sum_{j=1}^{n} \nu_{i,j}\left(\Gamma\mu_j - \sum_{h=1}^{d}\sum_{l=1}^{d} \nu_{h,l}\mu_l\right),$$

$$f_i(S_t,V_t,t) = -\left(\Gamma\mathcal{J}_{S_iV}S_i - \sum_{j=1}^{d} \mathcal{J}_{S_jV}S_j\sum_{h=1}^{d}\nu_{i,h}\right)\frac{1}{\Gamma V\mathcal{J}_{VV}}.$$

From previous expressions of C^* and w^*, we deduce the second-order differential equation satisfied by $\mathcal{J}$. Its coefficients are functions of S_t, V_t, and t:

$$0 = U(J(\mathcal{J}_V,t)) + \mathcal{J}_t + \mathcal{J}_V\left(\frac{\sum_{i=1}^{d}\sum_{i=1}^{d}\nu_{i,j}\mu_{i,t}V_t}{\Gamma} - J(\mathcal{J}_V,t)\right) \tag{6.51}$$

$$+\sum_{i=1}^{d}\mu_{i,t}S_{i,t}\mathcal{J}_{S_i} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_{i,j,t}S_{i,t}S_{j,t}\mathcal{J}_{S_iS_j} + \frac{V_t}{\Gamma}\sum_{j=1}^{d}S_{j,t}\mathcal{J}_{S_jV}$$

$$-\frac{\mathcal{J}_V}{\Gamma\mathcal{J}_{VV}}\left(\sum_{i=1}^{d}\Gamma\mu_{i,t}S_{i,t}\mathcal{J}_{S_iV} - \sum_{i=1}^{d}S_{i,t}\mathcal{J}_{S_iV}\sum_{j=1}^{d}\sum_{h=1}^{d}\nu_{,jl}\mu_{l,t}\right) + \frac{\mathcal{J}_{VV}V^2}{2\Gamma}$$
$$-\frac{1}{2\Gamma\mathcal{J}_{VV}}\left[\sum_{i=1}^{d}\sum_{j=1}^{d}\Gamma\sigma_{i,j,t}S_{i,t}S_{j,t}\mathcal{J}_{S_iV}\mathcal{J}_{S_jV} - \left(\sum_{i=1}^{d}S_{i,t}\mathcal{J}_{S_iV}\right)^2\right]$$
$$-\frac{\mathcal{J}_V^2}{2\Gamma\mathcal{J}_{VV}}\left[\sum_{i=1}^{d}\sum_{j=1}^{d}\Gamma\nu_{i,j,t}\mu_{i,t}\mu_{j,t} - \left(\sum_{i=1}^{d}\sum_{j=1}^{d}\nu_{i,j,t}\mu_{i,t}\right)^2\right].$$

In addition, the functional $\mathcal{J}$ satisfies the equation:

$$\mathcal{J}(V_T,S_T,T) = \widetilde{U}\left(V_T,T\right).$$

Then, the solution $\mathcal{J}^*$ of Equation (6.51) is introduced in equations (6.49) and (6.50), which allows us to deduce c^* and w^*.

For the general case, the solutions of these equations are not explicit. Moreover, for a large number of securities, numerical computations of solutions are tedious.

However, for some cases, explicit solutions can be identified and thus more easily analyzed; for example, if securities are defined from a Brownian multidimensional process with constant coefficients (in that case, they have lognormal distributions).

Assume that functions $\mu(.,.) = (\mu_1, ..., \mu_d)(.,.)$ and $\sigma(.,.) = (\sigma_1, ..., \sigma_d)(.,.)$ are constant. Equation (6.51) is:

$$0 = U(J(\mathcal{J}_V, t)) + \mathcal{J}_t + \mathcal{J}_V \left(\frac{\sum_{i=1}^{d} \sum_{i=1}^{d} \nu_{i,j} \mu_{i,t}}{\Gamma} V_t - J(\mathcal{J}_V, t) \right) \tag{6.52}$$

$$- \frac{\mathcal{J}_V^2}{2\Gamma \mathcal{J}_{VV}} \left[\sum_{i=1}^{d} \sum_{j=1}^{d} \Gamma \nu_{i,j,t} \mu_{i,t} \mu_{j,t} - \left(\sum_{i=1}^{d} \sum_{j=1}^{d} \nu_{i,j,t} \mu_{i,t} \right)^2 \right].$$

The solution $\mathcal{J}$ depends only on V, not on t or S. In that case, optimal weights w^* satisfy:

$$w^*_{i,t} = h_i + m(V_t, t) g_i, \tag{6.53}$$

where h_i and g_i are constant.

Therefore, at any time t, the weight $w^*_{i,t}$ belongs to a straight line included in the hyperplane $\sum_{i=1}^{d} w_i = 1$. The term $m(V_t, t)$ looks like a parameter which determines the position of $w^*_{i,t}$ on this line. It depends on the utility function, contrary to the terms h_i and g_i.

PROPOSITION 6.5

Under previous assumptions, the optimal solution w^* has the following decomposition: There exist two mutual-funds, independent from the investor's preferences, such that the investor is indifferent between a combination of these two funds and a combination of the d given securities. These two funds are characterized (up to a multiplicative coefficient) by their respective weighting

α and β such that for any $i \in \{1, ..., d\}$:

$$\alpha_i = h_i + \frac{1-\eta}{\nu} g_i, \tag{6.54}$$

$$\beta_i = h_i - \frac{\eta}{\nu} g_i, \tag{6.55}$$

where η and ν are arbitrary constant parameters ($\nu \neq 0$).

The solution w^* is such that there exists a scalar $a(V_t, t)$ satisfying

$$a(V_t, t) = \upsilon m(V_t, t) + \eta,$$

and

$$w^*_{i,t} = a(V_t, t)\alpha_i + [1 - a(V_t, t)]\, \beta_i. \tag{6.56}$$

If there exists a riskless asset S_0 with rate of return r, the proportions are given by: for any $i \in \{1, ..., d\}$,

$$\alpha_i = \frac{1-\eta}{\nu} \sum_{j=1}^{d} \nu_{ij}(\mu_j - r), \tag{6.57}$$

$$\beta_i = -\frac{\eta}{\nu} \sum_{j=1}^{d} \nu_{ij}(\mu_j - r), \tag{6.58}$$

$$\alpha_0 = 1 - \sum_{j=1}^{d} \alpha_i \text{ et } \beta_0 = 1 - \sum_{j=1}^{d} \beta_i. \tag{6.59}$$

REMARK 6.7 For a financial market with a return vector having a Gaussian distribution, two mutual funds are sufficient to generate all optimal portfolios. This dynamic two-fund separation property is analogous with the Tobin-Markowitz separation property, proved in the mean-variance framework (see Chapter 3). When a riskless asset exists, it can be considered as the first mutual fund. Note that the value of the second one also has a lognormal distribution.

□

When considering HARA utility functions, the optimal solution is an explicit function of the utility function parameters. Assume, to simplify, that there exist two securities, one riskless and one risky. Consider an intertemporal utility defined as a function of the consumption c of the following type:

$$U(c,t) = \exp(-\rho t) \times \widehat{U}(c) \text{ with } \widehat{U}(c) = \frac{1-\gamma}{\gamma} \left(\frac{\beta c}{1-\gamma} + \eta \right)^{\gamma}, \tag{6.60}$$

where $\rho > 0$, $\gamma \neq 1$, $\beta > 0$, $\frac{\beta C}{1-\gamma} + \eta > 0$, $\eta = 1$ if $\gamma = -\infty$.

The optimal solutions of Equation (6.52) are given by:

$$\mathcal{J}(V,t) = \frac{\delta\beta^{\gamma}}{\gamma}\exp(-\rho t)\left(\frac{\delta\left[1-\exp\left[-\left(\rho-\gamma\upsilon\right)\left(T-t\right)/\gamma\right]\right]}{\rho-\gamma\upsilon}\right)^{\delta} \tag{6.61}$$
$$\times\left(\frac{V}{\delta}+\frac{\eta}{\beta r}\left[1-\exp\left[-r(T-t)\right]\right]\right)^{\gamma}$$

where $\delta = 1-\gamma$ and $\nu = r + (\mu - r)^2/2\delta\sigma^2$.

Note that, for $\gamma > 1$, the solution holds only for

$$V_t \leq (\gamma-1)\eta\left[1-\exp\left[-r(T-t)\right]\right]/\beta r.$$

The optimal consumption and weighting are respectively given by:

$$c_t^*(V_t) = \frac{(\rho-\gamma\nu)\left(V_t+\frac{\delta\eta}{\beta r}\left[1-\exp\left[-r(T-t)\right]\right]\right)}{\delta\left(1-\exp\left[\frac{(\rho-\gamma\upsilon)(T-t)}{\delta}\right]\right)} - \frac{\delta\eta}{\beta}, \tag{6.62}$$

and,

$$w_t^*(V_t)V_t = \frac{\mu-r}{\delta\sigma^2}V_t + \frac{\eta\left(\alpha-r\right)}{\beta r\sigma^2}\left[1-\exp\left[-r(T-t)\right]\right]. \tag{6.63}$$

REMARK 6.8 The optimal consumption c^* and amounts w^*V are linear functions of the wealth V. The value process V also has coefficients which are deterministic functions w.r.t. the current time. Note that, when the vector of security logreturns is Gaussian, the HARA utility functions are the only utility functions for which these linearity properties are valid.

□

The final step is to search for the optimal wealth V^*. It is determined by using previous solutions c^* and w^* of Equation 6.62:

$$dV_t = \left(\left[w_t^*\left(\mu-r\right)+r\right]V_t - c_t^*\right)dt + \sigma w_t^* dW_t.$$

Then, setting

$$X_t = V_t + \frac{\delta\eta}{\beta r}\left[1-\exp\left[-r(T-t)\right]\right],$$

we have a stochastic process $(X_t)_t$ which is solution of the following standard SDE:

$$\frac{dX_t}{X_t} = adt + bW_t,$$

where a and b are constant. The solution X is a geometric Brownian motion. Then, its probability distribution is lognormal. We deduce the optimal

portfolio value V^*:

$$V_t^* = X_t^* - \frac{\delta\eta}{\beta r}\left[1 - \exp\left[-r(T-t)\right]\right], \tag{6.64}$$

with

$$X_t^* = X_0 \exp\left[\left(r - \frac{\rho - \gamma\nu}{\delta} + (1-2\gamma)\frac{(\mu - r)^2}{2\sigma^2\delta^2}\right)t + \frac{\mu - r}{\sigma\delta}W_t\right]$$
$$\times \frac{1 - \exp\left[\frac{\rho-\gamma\nu}{\delta}(t-T)\right]}{1 - \exp\left[-\frac{\rho-\gamma\nu}{\delta}(T)\right]}. \tag{6.65}$$

Then, the optimal consumption and allocations are functions $c_t^*(V_t^*)$ and $w_t^*(V_t^*)$.

Examine the following two particular cases:

- Case 1: (logarithmic utility) (which results from the limit case where $\gamma = \eta = 0$, and $\beta = 1 - \gamma = 1$)

$$c_t^*(V_t^*) = V_t^*/(T-t) \text{ and } w_{i,t}^* V_t^* = \left(\sum_{j=1}^{n} \nu_{ij}(\mu_j - r)\right) V_t^*. \tag{6.66}$$

- Case 2: (CARA utility) ($U(c,t) = -\exp(-\eta c)/\eta)$)

$$c_t^*(V_t^*) = rV_t^* + \frac{1}{\eta r}\left[\frac{(\mu - r)^2}{2\sigma^2} - r\right] \text{ and } w_t^* V_t^* = \frac{(\mu - r)}{\eta r \sigma^2}. \tag{6.67}$$

6.2.3 The general case

Consider the general financial model with all assumptions (A),(P), and (U). This model has been examined by Karatzas et al. [318], using the Bellman approach developed in the previous section. However, since the financial market is complete, another approach, based on the representation theorem for martingales, can be used to solve the optimization problem, as shown by Lehoczky *et al.* [350].

The investor searches for the solution of

$$\max_{C,w} E\left[\int_0^T U(C_s, s)ds + \widetilde{U}(V_T, T)\right], \tag{6.68}$$

under a positivity constraint on the wealth process V.

6.2.3.1 The positivity constraint

Indeed, the "infinite-dimensional" constraint $V_t \geq 0$ is not easy to check. Thus, we can try to "reduce" it to one-dimension. For this purpose, introduce the risk-neutral probability $\mathbb{Q}$ with Radon-Nikodym derivative L defined by:

$$L_t = \mathbb{E}\left[\frac{d\mathbb{Q}}{d\mathbb{P}} |\mathcal{F}_t\right] = \exp\left[\int_0^t \eta(s) dW_s - \frac{1}{2}\int_0^t ||\eta(s)||^2 \, ds\right], \tag{6.69}$$

where: ($\mathbb{I}$ denotes the vector with all components equal to 1)

$$\eta(t) = -\left[\sigma(t)\right]^{-1} \left(\mu(t) - r(t)\mathbb{I}\right). \tag{6.70}$$

Under assumptions (A), the process L is bounded and the Girsanov theorem can be used.

The process $\widetilde{W}$, defined by $\widetilde{W}_t = W_t - \int_0^t \eta(s) ds$, is a Brownian motion w.r.t. the risk-neutral probability $\mathbb{Q}$ and the filtration $\mathcal{F}_t$. Under $\mathbb{Q}$, the asset prices are solution of:

$$dS_t = S_{t_-}.\left(r_t dt + \sigma(t, S_t) d\widetilde{W}_t\right), \tag{6.71}$$

and the wealth process, associated to strategy (c, w), is given by:

$$dV_t^{c,w} = \left[V_{t-}^{c,w}\left(r_t - c(t)\right)\right] dt + \sum_{i,j} w_i(t)\sigma_{i,j}(t, S_t) d\widetilde{W}_{j,t}, \tag{6.72}$$

$$V^{c,w}(0) = v_0. \tag{6.73}$$

Denote by R the money market account:

$$R_t = \exp\left[-\int_0^t r(s) ds\right]. \tag{6.74}$$

Then

$$V_t^{c,w} R_t = v_0 - \int_0^t R(s)c(s)ds + \int_0^t R(s)^t w(s)\sigma(s) d\widetilde{W}_s.$$

DEFINITION 6.3 *A strategy (c, w) is said to be "admissible" for a given initial wealth v_0, if the process $V^{c,w}$ is positive a.s. Denote $\mathcal{A}(v_0)$ the set of such strategies.*

PROPOSITION 6.6
Let a strategy (c, w) in $\mathcal{A}(v_0)$ and $V^{c,w}$ represent the associated wealth process. Then

$$\mathbb{E}_{\mathbb{Q}}\left[V_T^{c,w} R_T + \int_0^T R(s)c(s)ds\right] \leq v_0. \tag{6.75}$$

PROOF If (c, w) is admissible, the process $M_t = v_0 + \int_0^t R(s)^t w(s)\sigma(s)d\widetilde{W}_s$ is a local $\mathbb{Q}$-martingale which is equal to $V_t^{c,w} R_t + \int_0^t R(s)c(s)ds$, and so is positive. Thus, M is a $\mathbb{Q}$-supermartingale, which implies $\mathbb{E}_{\mathbb{Q}}[M_T] \leq M_0$.

∎

A converse property can also be proved, using the martingale representation.

PROPOSITION 6.7

Let c be a consumption process satisfying assumption (S2), and let X be a positive random variable $\mathcal{F}_T$-measurable ("contingent claim") such that:

$$\mathbb{E}_{\mathbb{Q}}\left[X R_T + \int_0^T R(s)c(s)ds\right] = v_0.$$

Then, there exists a portfolio weighting w which is predictable such that the pair (c, w) is admissible, and the terminal wealth $V_T^{c,w}$ is equal to X.

PROOF

- First note that, if such weighting w exists, the local $\mathbb{Q}$-martingale

$$M_t = v_0 + \int_0^t R(s)^t w(s)\sigma(s)d\widetilde{W}_s, \tag{6.76}$$

is positive since it can be written as:

$$M_t = V_t^{c,w} R_t + \int_0^t R(s)c(s)ds.$$

The supermartingale M is also a martingale such that:

$$M_t = \mathbb{E}_{\mathbb{Q}}[M_T \,|\mathcal{F}_t] = \mathbb{E}_{\mathbb{Q}}\left[V_T^{c,w} R_T + \int_0^T R(s)c(s)ds \,|\mathcal{F}_t\right].$$

- Therefore, it is sufficient to prove that there exists w satisfying relation (6.76). For given pair (c, X), the process

$$M_t = \mathbb{E}_{\mathbb{Q}}\left[X R_T + \int_0^T R(s)c(s)ds \,|\mathcal{F}_t\right] \tag{6.77}$$

is a $\mathbb{Q}$-martingale.

The predictable representation theorem for martingales proves that there exists a predictable process θ such that $\int_0^T ||\theta_t||^2 \, dt < \infty$ a.s., and

$$M_t = M_0 + \int_0^t \theta(s)d\widetilde{W}_s \text{ with } M_0 = \mathbb{E}_{\mathbb{Q}}[M_T] = v_0. \tag{6.78}$$

Then, consider the process w given by: for any $t \in [0,T]$,

$$w(t) = -\left[R(t)\right]^{-1} \left[{}^{t}\sigma(t)\right]^{-1\ t}\theta(t),$$

and the process $V^{c,w}$ defined by:

$$V_t^{c,w} R_t = v_0 - \int_0^t R(s)c(s)ds + \int_0^t \theta(s)d\widetilde{W}_s.$$

We can easily check that the process $V^{c,w}$ is associated to the strategy (c,w), which is admissible.

□

6.2.3.2 Existence of an optimal strategy

The optimization problem is:

$$\max_{c,w}\ \mathbb{E}_{\mathbb{P}}\left[\int_0^T U(c_s,s)ds + \widetilde{U}\left(V_T,T\right)\right], \tag{6.79}$$

with the budget constraint:

$$\mathbb{E}_{\mathbb{P}}\left[V_T^{c,w}R(T)L(T) + \int_0^T R(s)c(s)L(s)ds\,|\mathcal{F}_t\right] \leq v_0.$$

Consider the value-function $\mathcal{J}$:

$$\mathcal{J}(V^{c,w},w,t,v_0) = \max_{c,w}\ \mathbb{E}_{\mathbb{P}}\left[\int_t^T U(c_s,s)ds + \widetilde{U}\left(V_T^{c,w},T\right)|\mathcal{F}_t\right]. \tag{6.80}$$

We first have to search for optimal solution (c^*,V^*), then to determine w^*. A useful duality result can be used.

PROPOSITION 6.8 Legendre transform

Under the assumptions (U), we have:

$$U(J(y)) - yJ(y) = \max_{c\geq 0}\left[U(c) - cy\right],$$

where J denotes the inverse function of U'. Note that

$$\widehat{U}(y) = \max_{c\geq 0}\left[U(c) - cy\right]$$

is the convex conjugate function of the investor's utility.

PROOF Since U is concave, we have:

$$U(J(y)) - U(c) \geq U'(J(y))(J(y)) - c).$$

If $J(y) > 0$, $U'(J(y)) = y$. Therefore:

$$U(J(y)) - U(c) \geq y(J(y)) - c).$$

If $J(y) = 0$, $y \geq U'(0)$. Therefore:

$$U(0) - U(c) \geq -U'(0)c \geq -yc.$$

∎

To determine an optimal solution, we can use the Lagrange multipliers. For $\lambda \in \mathbb{R}^+$, denote:

$$\mathcal{L}(c, V_T, \lambda) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T U(c_s, s)ds + \widetilde{U}\,(V_T, T)\right]$$

$$+\lambda\left(v_0 - \mathbb{E}_{\mathbb{P}}\left[V_T R(T) L(T) + \int_0^T R(t)c(t)L(t)dt\right]\right).$$

A sufficient condition for (c^*, V^*) to be optimal is that there exists a Lagrange multiplier $\lambda^* \in \mathbb{R}^+$ such that (c^*, V_T^*, λ^*) satisfies: for any (c, V_T, λ) satisfying budget equation 6.75 and $\lambda \in \mathbb{R}^+$,

$$\mathcal{L}(c, V_T, \lambda^*) \leq \mathcal{L}(c^*, V_T^*, \lambda^*) \leq \mathcal{L}(c^*, V_T^*, \lambda).$$

The first inequality, which is due to the optimality of (c^*, V_T^*), can be solved by searching for the values (c_t^*, V_T^*) which satisfy: for any (t, ω),

- The consumption c_t^*maximizes $U(c_t(\omega), t) - \lambda^* R(t) c_t(\omega) L(t)$; and,

- The wealth value V_T^* maximizes $\widetilde{U}(V_T^*(\omega), T) - \lambda^* R(T) V_T^*(\omega) L(T)$.

We deduce:

$$c_t^* = J(\lambda^*(\kappa(t))) \text{ and } V_T^* = \widetilde{J}(\lambda^*(\kappa(T))),$$

where $\kappa(t) = R(t)L(t)$.

The Lagrange parameter λ^* is determined from the budget equation. Its existence is deduced by assuming:

- (V): For any $\lambda > 0$, $\mathbb{E}\left[\int_0^T L(t)R(t)J(\lambda\kappa(t))dt\right]$ and $\mathbb{E}\left[L(T)R(T)\widetilde{J}(\lambda\kappa(T))\right]$ are finite.

Under assumptions (U) and (V), the function F, defined by

$$F(y) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T \kappa(t)J(y\kappa(t),t)dt + \kappa(T)\widetilde{J}\left(y\kappa(T),T\right)\right],$$

is continuous and non-increasing on $[0,a]$ where $a = \inf\{y\,|F(y)=0\}$. It satisfies also:

$$\lim_{y\to 0} F(y) = +\infty; \quad \lim_{y\to+\infty} F(y) = 0.$$

Then, the function F has an inverse F^{-1}. Define the function G by:

$$G(y) = H(F^{-1}(y)),$$

where the function H is given by:

$$H(y) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T U(J(y\kappa(t)),t)dt + \widetilde{U}\left[\widetilde{J}\left(y\kappa(T),T\right)\right]\right].$$

Assume also:

- (W): the utility functions U and $\widetilde{U}$ are twice-continuously differentiable and U" and $\widetilde{U}$" are increasing.

Then (see Karatzas *et al.* [318]): F and G are continuously-differentiable and $H'(y) = yF'(y)$.

To summarize:

PROPOSITION 6.9

Under assumptions (U),(V), and (W) for utility functions U and $\widetilde{U}$, there exists an optimal strategy $(c^*,w^*)\in\mathcal{A}(v_0)$ such that:

$$\mathcal{J}(c^*,w^*,v_0) = \max_{c,w\in\mathcal{A}(v_0)} \mathcal{J}(c,w,v_0) \tag{6.81}$$

$$\text{where } \mathcal{J}(c,w,v_0) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T U(c_s,s)ds + \widetilde{U}\left(V_T^{c,w},T\right)\right]. \tag{6.82}$$

Using previous function F and the inverse functions J and $\widetilde{J}$ of marginal utility functions U' and $\widetilde{U}'$, we have:

$$c_t^* = J(F^{-1}(v_0)\kappa(t)), \tag{6.83}$$

$$V_T^* = \widetilde{J}(F^{-1}(v_0)\kappa(T)), \tag{6.84}$$

and the optimal weighting w^* is deduced from the martingale representation (6.78).

Example 6.5

Assuming that the coefficients r, μ, and σ are constant, that the utility function U is a power function, $U(x) = \frac{x^\gamma}{\gamma}$, and that $\widetilde{U}(x) = 0$. We have:

1) The inverse function J is given by: $J(y) = \left(\frac{y}{\gamma}\right)^{1/(\gamma-1)}$.

2) The function F^{-1} is such that:

$$F^{-1}(v) = \gamma\,(bv)^{(\gamma-1)} \text{ with } b = 1/\mathbb{E}\left[\int_0^T \kappa(t)^{\gamma/(\gamma-1)}dt\right].$$

3) The optimal consumption is given by:

$$c_t^* = J(F^{-1}(v_0)\kappa(t)) = v_0 b\kappa(t)^{1/(\gamma-1)}.$$

4) The optimal wealth is null, since the utility function $\widetilde{U}$ is itself null.

5) The optimal weighting is determined from the predictable representation of the martingale

$$M_t = \mathbb{E}_{\mathbb{Q}}\left[\int_0^T R(s)c^*(s)\,|\mathcal{F}_t\right] = v_0 + \int_0^t \theta^*(s)d\widetilde{W}_s\ ,$$

which, by identification, leads to the equality:

$$w^*(t) = [R(t)]^{-1}\,\theta^*(t)\,[\sigma(t)]^{-1} = e^{rt}\sigma^{-1}\theta^*(t).$$

To determine $\theta^*(t)$, we note that:

$$\kappa(t) = e^{-rt}\exp\left[\eta W_t - \eta^2 t/2\right].$$

Then, setting

$$\delta = 1/\left(\gamma - 1\right),$$

consider the exponential martingale:

$$\widetilde{L}_t = \exp\left[\eta\delta\widetilde{W}_t - \eta^2\delta^2 t/2\right].$$

Set

$$\zeta = r(1+\delta) - \frac{1}{2}\eta^2\delta(1+\delta).$$

Then,

$$R_t c^*(t) = v_0 b R_t \kappa(t)^\delta = v_0 b\widetilde{L}_t e^{-\zeta t}.$$

Thus, the martingale M_t can be expressed by using $\widetilde{L}$:

$$M_t = v_0 b \int_0^t \widetilde{L}_s e^{-\zeta s} ds + v_0 b \int_t^T \mathbb{E}_{\mathbb{Q}} \left[\widetilde{L}_s \,|\mathcal{F}_t \right] e^{-\zeta s} ds.$$

Since $\widetilde{L}$ is a $\mathcal{F}_t$-martingale w.r.t. $\mathbb{Q}$, we have:

$$M_t = v_0 b \left[\int_0^t \widetilde{L}_s e^{-\zeta s} ds + \frac{e^{-\zeta t} - e^{-\zeta T}}{\zeta} \widetilde{L}_t \right].$$

Using $d\widetilde{L}_t = \eta b \widetilde{L}_t d\widetilde{W}_t$, we deduce:

$$dM_t = \frac{v_0 b \eta \delta}{\zeta} \left[\widetilde{L}_t \left(e^{-\zeta t} - e^{-\zeta T} \right) \right] d\widetilde{W}_t.$$

Finally,

$$w^*(t) = \frac{v_0 b(\mu - r)}{\sigma^2 (\gamma - 1)\zeta} \left[\left(e^{-\zeta(T-t)} - 1 \right) \right] \kappa(t)^{1/(\gamma-1)},$$

and

$$H(y) = \left(\frac{y}{\gamma} \right)^{\gamma/(\gamma-1)} \frac{1 - e^{-\zeta T}}{\zeta},$$
$$G(v) = (v)^{\gamma} \left(\frac{1 - e^{-\zeta T}}{\zeta} \right)^{(1-\gamma)}.$$

Note that if $\widetilde{U}(x) = U(x) = \frac{x^\gamma}{\gamma}$, we have:

$$c_t^* = \frac{v_0}{\widehat{b}} \left(\frac{\kappa(t)}{\gamma} \right)^{1/(\gamma-1)} \text{and } V_T^* = \frac{v_0}{\widehat{b}} \left(\frac{\kappa(T)}{\gamma} \right)^{1/(\gamma-1)},$$

with

$$\widehat{b} = 1/\mathbb{E} \left[\int_0^T \kappa(t)^{\gamma/(\gamma-1)} dt + \kappa(T)^{\gamma/(\gamma-1)} \right] \gamma^{1/(\gamma-1)}.$$

□

REMARK 6.9 The previous method relies heavily on martingale methods. In Chapter 7, this approach is used to determine optimal portfolio design, in particular by using the results of Cox and Huang [132].

□

6.2.4 Recursive utility in continuous-time

Epstein and Zin [212] introduce the notion of recursive preferences which generalizes the standard time-separable power utility. This new preference modelling allows for the separation of the relative risk aversion from the elasticity of intertemporal substitution of consumption. Duffie and Epstein [173] (see also Svensson [483]) introduce recursive utility in continuous-time.

Consider the following parametrization of recursive utility:

$$U_{re,t} = \mathbb{E}\left[\int_t^{\infty} f\left(C_s, U_{re,s}\right) ds \,|\mathcal{F}_t\right], \tag{6.85}$$

where $f(.,.)$ is a function which allows for the aggregation of the current consumption C_s and the continuation utility I_s.

Duffie and Epstein [173] assume that this function is given by:

$$f(C,J) = \frac{\beta}{1-\left(\frac{1}{\psi}\right)}\left(1-\gamma\right)U_{re}\left(\left[\frac{C}{\left(\left(1-\gamma\right)U_{re}\right)^{\frac{1}{1-\gamma}}}\right]^{1-\frac{1}{\psi}} - 1\right), \tag{6.86}$$

where $\beta > 0$ denotes the rate of time preference, $\gamma > 0$ denotes the coefficient of relative rsik aversion, and $\psi > 0$ denotes the elasticity of intertemporal substitution.

Note that for $\psi = \frac{1}{\gamma}$, we recover the power utility. Duffie and Epstein [173] prove that:

- First, the Bellman principle can still be applied.
- Second, it is sufficient to substitute the term $f(C, U_{re})$ for the instantaneous utility function $U(C)$in Equation (6.43). Then, we have to use a new functional Φ_r defined by:

$$\Phi_r(C, w, V, S, t) = f(C, U_{re}) + \mathcal{L}(\mathcal{J}). \tag{6.87}$$

Example 6.6

In the recursive utily framework, Campbell and Viceira [102] propose to examine the impact of a mean-reverting interest-rate. They assume that the investor can only choose between cash and a long-term real bond. They also assume that the instantaneous riskless interest rate r_t follows an Ornstein-Uhlenbeck process given by:

$$dr_t = a_r(b_r - r_t)dt - \sigma_r dW_r, \tag{6.88}$$

where a_r, b_r, and σ_r are positive constants, and W_r is a standard Brownian motion. The market price of interest rate risk is assumed to be constant (see Vasicek [498]).

(1) Cash dynamics:

$$\frac{dC_t}{C_t} = r_t dt. \tag{6.89}$$

(2) The zero-coupon bond B with maturity T is solution of the following SDE:

$$\frac{dB(r_t, T-t)}{B(r_t, T-t)} = \left[r_t + \theta(T-t)\right] dt + \sigma(T-t) dW_{r,t}, \tag{6.90}$$

where σ_{T-t} denotes the volatility at time t of the zero-coupon bond:

$$\sigma_{T-t} = \frac{\sigma_r(1 - e^{-a_r(T-t)})}{a_r}, \tag{6.91}$$

which is decreasing with time.

Denote by λ_r the risk premium on the bond. Set

$$\theta(T-t) = \lambda_r \sigma_{T-t}.$$

Using the Bellman equation associated to the functional Φ_r in Equation (6.87) with $\psi = 1$, Campbell and Viceira guess that the solution has the following form:

$$U_{re}(V_t, r_t) = I(r_t) \frac{V_t^{1-\gamma}}{1-\gamma}. \tag{6.92}$$

This leads to the following ordinary differential equation (ODE):

$$\begin{aligned} 0 = &\frac{\beta}{1-\gamma} \log(I) + \left(\beta \log \beta - \beta + \frac{\lambda_r^2}{2\gamma} + r\right) + \frac{\sigma_r^2}{2\gamma} \left(\frac{1}{I}\frac{\partial I}{\partial r}\right)^2 \\ &+ \left(\frac{a_r}{1-\gamma}(b_r - r) - \frac{\lambda_r \sigma_r}{\gamma}\right) \frac{1}{I}\frac{\partial I}{\partial r} + \frac{\sigma_r^2}{2(1-\gamma)} \frac{1}{I}\frac{\partial^2 I}{\partial r^2}. \end{aligned} \tag{6.93}$$

This equation has an exact solution. There exist two constants A_0 and A_1 such that:

$$I(r_t) = \exp\left[A_0 + A_1 r_t\right].$$

The consumption-wealth ratio is constant equal to β, and the fraction of wealth α_t invested on the bond is given by:

$$\alpha_t = \frac{1}{\gamma}\frac{\lambda_r}{\sigma_{T-t}} + \left(1 - \frac{1}{\gamma}\right) \frac{\sigma_r}{\sigma_{T-t}(a_r + \beta)}.$$

□

6.3 Further reading

Optimization theory and its applications based on ordinary equations are studied in Cesari [110]. Multiperiod optimization based on discrete-time Bellman principle is examined in Bertsekas ([63], [64]). Portfolio optimization in the binomial model is solved in Bajeux [37].

Since this chapter provides only a brief overview on stochastic control theory, it refers to El Karoui [188] for a general and sound stochastic control analysis. Results about controlled diffusion process can be found in Krylov [339], Davis [150], and Dempster [158].

Optimization dealing with Markov models are detailed in [150]. Hamilton-Jacobi equations can be studied by using viscosity notion, as shown for example by Shreve and Soner [470], and Zariphopoulou [510].

Dynamic allocation problems are also examined in El Karoui and Karatzas [189]. Recursive portfolio analysis is provided in El Karoui, Peng and Quenez [195], using backward stochastic differential equations.

Portfolio optimization in a lognormal market is examined for a power utility in Dexter *et al.* [166].

Portfolio optimization when assets are discontinuous is examined in Jeanblanc and Pontier [298], and in Shirakawa [468].

Numerical methods can be found in Kushner and Dupuis [341], Fitzpatrick and Fleming [229], and Rogers and Talay [429].

Monte Carlo methods for portfolio analysis are used in Cvitanic *et al.* [136], and in Detemple *et al.* [164].

Chapter 7

Optimal payoff profiles and long-term management

This chapter deals with two main applications of some of the results obtained in Chapter 6:

- First, the optimal portfolio value at maturity is assumed to be a function of a given set of asset prices. Therefore, the problem consists in determining an optimal function or "payoff profile."
- Second, the determination of an optimal long-term portfolio, invested in cash, bonds, and stocks.

7.1 Optimal payoffs as functions of a benchmark

7.1.1 Linear versus option-based strategy

Assume that the investor maximizes his expected utility, but uses only a buy-and-hold strategy $w = (w_{B,0}, w_{S,0})$ to invest in a riskless asset B with rate r, and a risky asset S. Then, any solution of the optimization problem:

$$Max_w \mathbb{E}_{\mathbb{P}}[U(V_T)] \text{ with } V_0 = e^{-rT}\mathbb{E}_{\mathbb{Q}}V_T],$$

is necessarily linear w.r.t. the risky asset:

$$V_T = V_0 \left(w_{B,0}\frac{B_T}{B_0} + w_{S,0}\frac{S_T}{S_0} \right).$$

However, he may search for nonlinear solutions, including, for example, derivatives in his portfolio.

To determine his optimal choice, V_T^*, he can search this solution assuming that $V_T = h(S_T)$. Then, the portfolio optimization is now given by:

$$Max_h \mathbb{E}_{\mathbb{P}}[U(h(S_T))] \text{ with } V_0 = e^{-rT}\mathbb{E}_{\mathbb{Q}}[h(S_T)],$$

where $\mathbb{P}$ denotes the historical probability, and $\mathbb{Q}$ is the risk-neutral probability which is used to price options. According to Brennan and Solanki [88] (see

also Carr and Madan [106]), we deduce:

$$V_T^* = J(\lambda g), \tag{7.1}$$

where g is the pdf of $d\mathbb{Q}/d\mathbb{P}$ w.r.t. $\mathbb{P}$, λ is the Lagrange parameter corresponding to the budget constraint, and $J(x) = (U')^{-1}(x)$.

7.1.1.1 The general model

Suppose, as in [415], that three basic financial assets are available:

- The cash associated to a discount factor N.
- The bond B.
- The stock or a financial index S. The investor is supposed to determine an optimal payoff h which is a function defined on all possible values of the assets (N, B, S) at maturity.

REMARK 7.1 If the market is complete, this payoff can be achieved by the investor. The market can be complete, for example, as shown in the previous chapter if:

- The financial market evolves in continuous time and all options can be dynamically replicated by a perfect hedging strategy;
- Or, if, in one period setting, European options of all strikes are available on the financial market. In this setting, the inability to continuously trade potentially induces investment in cash, asset B, asset S and all European options with underlying assets B and S (if cash and bond are non stochastic, only European options on S are required).

The market can be also incomplete. In that case, the solution given in this section is only "theoretical," but still interesting to know since the optimal payoff can be approximated by investing on traded assets. (In practice, the investor defines an approximation method, which may take transaction costs or liquidity problems into account.)

$\square$

The investor is assumed to be a pricetaker. For example, if his benchmark S is the SP&500 then his investment is too weak to modify the index value.

Under the standard condition of no-arbitrage, the assets prices are calculated under risk neutral probabilities. If markets exist for out-of-the-money European puts and calls of all strikes, then it implies the existence of a unique risk-neutral probability that may be identified from option prices. Otherwise, if there is no continuous trading, generally the market is incomplete and one

particular risk-neutral probability $\mathbb{Q}$ is used to price the options. It is also possible that stock prices change continuously, but the market may be still dynamically incomplete. Again, it is assumed that one risk-neutral probability is selected. Assume that prices are determined under measure $\mathbb{Q}$. Denote by $\frac{d\mathbb{Q}}{d\mathbb{P}}$ the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to the historical probability $\mathbb{P}$. Denote by N_T the discount factor and by M_T the product $N_T \frac{d\mathbb{Q}}{d\mathbb{P}}$.

Due to the no-arbitrage condition, the budget constraint corresponds to the following relation:

$$V_0 = \mathbb{E}_{\mathbb{Q}}[h(N_T, B_T, S_T)N_T] = \mathbb{E}_{\mathbb{P}}[h(N_T, B_T, S_T)M_T].$$

The investor has to solve the following optimization problem:

$$Max_h \mathbb{E}_{\mathbb{P}}[U(h(N_T, B_T, S_T)] \text{ under } V_0 = \mathbb{E}_{\mathbb{P}}[h(N_T, B_T, S_T)M_T]. \tag{7.2}$$

To simplify the presentation of the main results, we suppose that the function h fulfils:

$$\int_{\mathbb{R}^{+3}} h^2(n, b, s)\mathbb{P}_{(N_T,B_T,S_T)}(dn, db, ds) < \infty \ .$$

This means that $h \in \mathbb{L}^2(\mathbb{R}^{+3}, \mathbb{P}_{X_T}(dx))$, where $X_T = (N_T, B_T, S_T)$, which is the set of the measurable functions with squares that are integrable on $\mathbb{R}^{+3}$ with respect to the distribution $\mathbb{P}_{X_T}(dx)$.

The utility function U is associated with a new functional Φ_U, which is defined on the space $\mathbb{L}^2(\mathbb{R}^{+3}, \mathbb{P}_{X_T}(dx))$ by:

$$\text{For any } Y \in \mathbb{L}^2(\mathbb{R}^{+3}, \mathbb{P}_{X_T}(dx)),\ \ \Phi_U(Y) = \mathbb{E}_{\mathbb{P}_{X_T}}[U(Y)].$$

Φ_U is usually called the Nemitski functional associated with U (see Ekeland and Turnbull [187] for definition and basic properties).

PROPOSITION 7.1

Introduce the conditional expectation of M_T under the σ-algebra generated by (N_T, B_T, S_T). Denote its pdf by g. Assume that g is a function defined on the set of the values of $X_T = (N_T, B_T, S_T)$, and $g \in \mathbb{L}^2(\mathbb{R}^{+3}, \mathbb{P}_{X_T})$. Then, the optimization problem is reduced to:

$$Max_{h \in \mathbb{L}^2(\mathbb{R}^{+3}, \mathbb{P}_{X_T})} \int_{\mathbb{R}^{+3}} [U(h(x))]\mathbb{P}_{X_T}(dx) \tag{7.3}$$

$$under: V_0 = \int_{\mathbb{R}^{+3}} h(x)g(x)\mathbb{P}_{X_T}(dx).$$

We deduce that the optimal payoff h^e is given by:

$$h^e = J(\lambda g), \tag{7.4}$$

where λ is the scalar Lagrange multiplier such that

$$V_0 = \int_{\mathbb{R}^{+3}} J(\lambda g(x))g(x)\mathbb{P}_{X_T}(dx).$$

PROOF From the properties of the utility function U, the Nemitski functional Φ_U is concave and differentiable (the Gâteaux-derivative exists) on $\mathbb{L}^2(\mathbb{R}^{+3}, \mathbb{P}_{X_T})$. Additionally, the budget constraint is a linear function of h. So there exists exactly one solution h^e.
h^e is the solution of $\frac{\partial L}{\partial h} = 0$ where the Lagrangian L is defined by:

$$L(h, \lambda) = \int_{\mathbb{R}^{+3}} [U(h(x))]\mathbb{P}_{X_T}(dx) + \lambda \left(V_0 - \int_{\mathbb{R}^{+3}} h(x)g(x)\mathbb{P}_{X_T}(dx) \right),$$

where λ is the Lagrange multiplier associated to the budget constraint.
So, h^e satisfies: $U'(h^e) = \lambda g$. Therefore, $h^e = J(\lambda g)$.

∎

Suppose for example that cash and bonds are not stochastic. Then, the properties of the optimal payoff h^* as a function of the benchmark S can be analyzed. Since the utility function U is concave, the marginal utility U' is decreasing, then J also is decreasing, from which we deduce:

COROLLARY 7.1
h^* is an increasing function of the benchmark S_T iff the conditional expectation g of $\frac{d\mathbb{Q}}{d\mathbb{P}}$ under the σ-algebra generated by S_T is a decreasing function of S_T. More precisely: assume that g is differentiable. From the optimality conditions, the derivative of the optimal payoff is given by:

$$h'(s) = \left(-\frac{U'(h(s))}{U''(h(s))} \right) \times \left(-\frac{g(s)'}{g(s)} \right).$$

Note that, in most cases, g is decreasing.

Introduce the tolerance of risk $T_o(h(s))$ equal to the inverse of the absolute risk-aversion:

$$T_o(h(s)) = -\frac{U'(h(s))}{U''(h(s))}.$$

As it can be seen, $h'(s)$ depends on the tolerance of risk. The design of the optimal payoff can also be specified. Denote $Y(s) = -\frac{g'(s)}{g(s)}$.

Differentiating twice with respect to s, and from the previous corollary, we deduce:

COROLLARY 7.2
Assume that g is twice-differentiable.
Then:

$$h''(s) = [X'(h(s)) + \frac{Y'(s)}{Y(s)^2}] \times [X(h(s))Y^2(s)]. \tag{7.5}$$

Therefore, usually, the higher the tolerance of risk, the higher $h''(s)$.

Example 7.1
In what follows, the optimal portfolio profile is examined for a special case.

- The utility function is CRRA:

$$U(x) = \frac{x^\alpha}{\alpha}, 0 < \alpha < 1,$$
$$J(x) = (U'(x))^{-1} = x^{\frac{1}{\alpha-1}}.$$

- The stock price S is a geometric Brownian motion. At each time t, its logarithm has a Gaussian probability distribution with mean $\left(\mu - \frac{1}{2}\sigma^2\right) t$ and variance $\sigma^2 t$.

The optimal payoff profile is given by:

$$h^*(s) = \frac{V_0 e^{rT}}{\int_0^\infty g(s)^{\frac{\alpha}{\alpha-1}} f_{\mathbb{P}}(s)ds} g(s)^{\frac{1}{\alpha-1}},$$

where s denotes all possible values of S_T, and $f_{\mathbb{P}}$ is the pdf of S_T:

$$l(S,\mu,\sigma) = \frac{\mathbf{I}_{\{x>0\}}}{S\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\ln(S/S_0) - \left(\mu - \frac{1}{2}\sigma^2\right)}{\sigma}\right]^2}.$$

Within this framework, the density g of the Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$ is given by:

$$g(s) = \psi s^{-\kappa},$$

with $\theta = \frac{\mu - r}{\sigma}$, $A = -\frac{1}{2}\theta^2 T + \frac{\theta}{\sigma}\left(\mu - \frac{1}{2}\sigma^2\right) T$, $\kappa = \frac{\theta}{\sigma}$ and $\psi = e^A (S_0)^\kappa$.

Therefore $h^*(s)$ can be written as a power function of the stock S:

$$h^*(s) = d.s^m,$$

with

$$d = \frac{V_0 e^{rT}}{\int_0^\infty g(s)^{\frac{\alpha}{\alpha-1}} f_{\mathbb{P}}(s)ds} \psi^{\frac{1}{\alpha-1}} > 0 \text{ and } m = \frac{\kappa}{1-\alpha} > 0.$$

Note that $h^*(s)$ is an increasing function of s. Its profile only depends on the comparison between the relative risk aversion $1-\alpha$ and the ratio κ, which looks like a Sharpe ratio, depending only on μ, r, and σ:

- $h^*(s)$ is concave if $\kappa < 1 - \alpha$.

  This means that, for a bearish market, the investor wants to receive higher payoff than the stock value. However, he will get a smaller payoff if the financial market will be bullish.

- $h^*(s)$ is linear if $\kappa = 1 - \alpha$.

  The investor buys the stock S itself.

- $h^*(s)$ is convex if $\kappa > 1 - \alpha$.

  The investor has a higher risk exposure in order to benefit better from a bullish market. He is less protected against a bearish market.

The following figure illustrates the concavity/convexity of the optimal portfolio profiles according to the risk-tolerance:

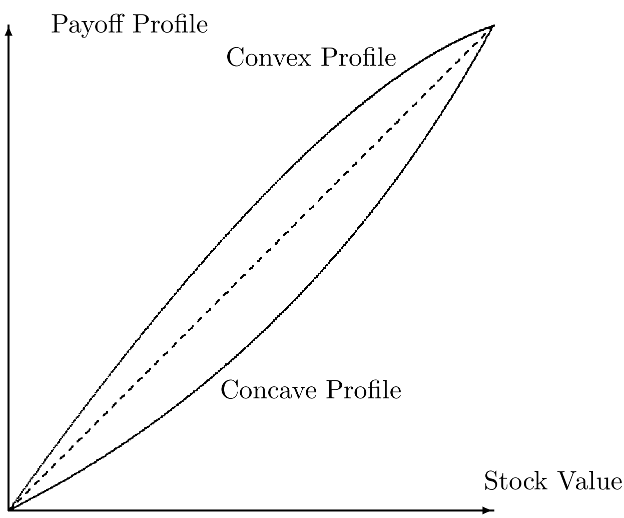


FIGURE 7.1: Optimal portfolio profiles

For a fixed relative risk aversion α, the influence of the market parameters can be examined: for example, the instantaneous expected return μ which is a fundamental parameter when dealing with portfolio management. As shown in next figure, the higher the parameter μ, the more convex the optimal portfolio payoff. Note also that, for "extreme" values of μ smaller than the riskless rate r, the optimal profile is a decreasing function of the stock price at maturity.

$\square$

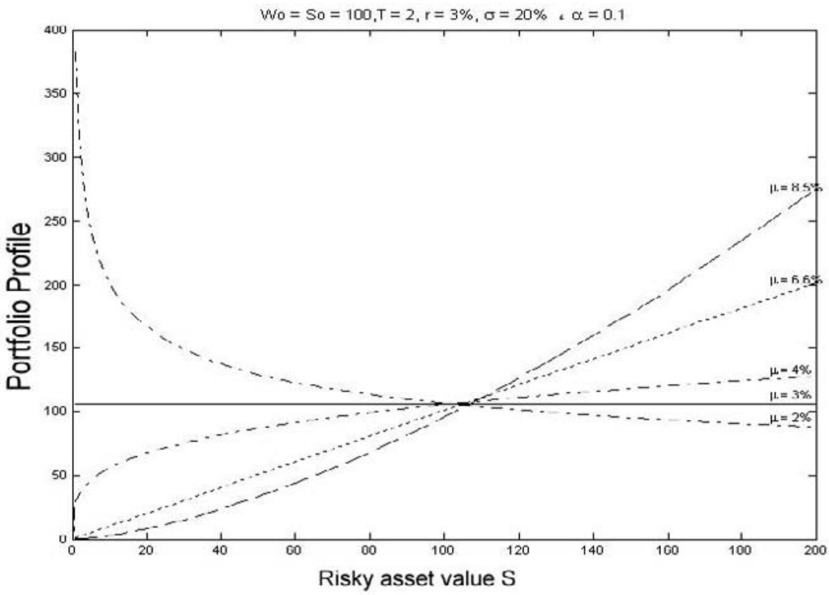


FIGURE 7.2: Optimal portfolio profiles according to stock return

REMARK 7.2 For an HARA utility function:

$$U(x) = a\left(b + \frac{x}{c}\right)^{1-c}, 0 < c < 1$$

$$J(x) = (U'(x))^{-1} = c\left[\left(\frac{cx}{a(1-c)}\right)^{-\frac{1}{c}} - b\right].$$

Then, the optimal portfolio profile has the following form:

$$V_T^* = h^*(S_T), \text{ with } h^*(s) = c\left[\left(\frac{cs^{-\kappa}}{a(1-c)}\right)^{-\frac{1}{c}} - b\right].$$

Consequently, the optimal portfolio profile is a linear function of a power of the stock value.

- This power is equal to $\frac{\kappa}{c}$. Note that c is the asymptotic relative risk aversion ($c = lim_{x\to\infty}\frac{-xU"(x)}{'(x)}$). Therefore, the discussion about concavity/convexity is the same as the previous one when dealing with CRRA utility functions.

- The term $-(bc)$ corresponds to a fixed guaranteed amount if it is positive, or a maximal loss if it is negative.

$\square$

7.2 Application to long-term management

Consider an investor who has a specific goal such as retirement, paying for his children's education, *etc.* He is facing a variety of decisions:

Which amount of money should he invest initially? How should he invest among assets: cash, stock index, and bond funds? Should he use a market timing or fixed strategy? For example, Dybvig [182] has shown that the cost of undiversified strategies over time can be substantial.

As mentioned in Chapter 1, a well-known property of an optimal portfolio is the *mutual fund separation* theorem: a rational investor divides his investment between two assets, a riskless one and a risky mutual fund, the composition of which is the same, whatever the investor's risk aversion.

As noted by Canner, Mankiw, and Weil [104], popular investment advice does not conform to this property. Empirical studies show that allocations between stocks, bonds and cash depend indeed on risk aversion. In particular, the ratios of bond/stock differ when considering conservative, moderate, or aggressive investors. For example, this ratio is 1.5 for a conservative investor, 1.00 for a moderate one, and 0.5 for an aggressive one.

Bajeux-Besnainou, Jordan, and Portait [38] address this inconsistency issue between mutual fund property and popular advice. They consider that the investor's horizon exceeds the maturity of the cash asset. They introduce a continuous-time portfolio rebalancing. In that case, cash may be a money market security with a short maturity (one to six months), and may no longer be the common riskless asset in the standard theory. In particular, this is true when dealing with long-term investments. Nevertheless, the investor can synthesize a riskless asset (for example a zero-coupon bond maturing at the horizon) by using a bond fund and cash. Consequently, bonds appear in both synthetic riskless asset and in the risky mutual fund, which can justify a bond/stock ratio varying with risk aversion for any HARA investor.

7.2.1 Assets dynamics and optimal portfolios

We adopt the same framework. This is a generalization of the Black and Scholes model and a variant of the Merton (1971) one state variable model. The model assumes normality of log returns, which is not a restrictive assumption when dealing with long term investment. In particular, this allows us to provide explicit formulas, which greatly simplifies the computation of utility and monetary losses. Obviously, other financial market models can be introduced and examined, as seen in Chapter 8.

7.2.1.1 The financial market

The market is assumed to be arbitrage-free and without friction. Financial transactions occur in continuous-time, along a time period $[0, T]$.

Three basic assets are available at any time on the market. (1) An instantaneously riskless *money market* fund, the *Cash,* with a price denoted by C. (2) A *Stock* index fund with a price S. (3) A *Bond* fund with constant duration D, obtained by continuously rolling bonds throughout the investment period $[0, T]$. It is denoted by B_D, which is a zero-coupon bond with maturity $(t+D)$ at time t.

As mentioned in [38], if inflation uncertainty is ignored, the interest rate risk means real estate rate risk. The omission of inflation uncertainty may induce some problems, when considering long horizons. Nevertheless, first empirical evidence indicates, for example, that the US inflation volatility is smaller than real interest rate volatility. Second, special long term bonds indexed on inflation are now available on some financial markets (for example, in the US or UK). Finally, it is possible to diversify by investing in the housing market.

Thus, since $D > d$, the riskless asset is a zero-coupon nominal bond B_T that matures at the investor's horizon, and which is replicated by a dynamic combination of C, B_D and S. Therefore, there exists a two-fund Cass-Stiglitz separation with a synthetic riskless fund B_T replicated by C, and B_D and a risky fund replicated by C, B_D, and S.

Since continuous-time rebalancing is allowed, financial markets can be assumed to be complete by introducing two sources of risk. In fact, as shown in Duffie and Huang [175], such assumption of dynamic market completeness is allowed when contingent claims can be synthesized by continuous-time rebalancing.

To illustrate the results, we assume that the instantaneous riskless interest rate r_t follows an Ornstein-Uhlenbeck process given by:

$$dr_t = a_r(b_r - r_t)dt - \sigma_r dW_r, \tag{7.6}$$

where a_r, b_r, and σ_r are positive constants, and W_r is a standard Brownian motion. The market price of interest rate risk is assumed to be constant (see Vasicek [498]).

The asset price dynamics are given by:

(1) Cash:

$$\frac{dC_t}{C_t} = r_t dt. \tag{7.7}$$

(2) Stock index:

$$\frac{dS_t}{S_t} = (r_t + \theta_S)dt + \sigma_1 dW + \sigma_2 dW_r. \tag{7.8}$$

(3) Bond fund:

$$\frac{dB_{Dt}}{B_{Dt}} = (r_t + \theta_B)dt + \sigma_B dW_r, \tag{7.9}$$

where W is another standard Brownian motion, independent of W_r, and where σ_1, σ_2, and σ_B are positive constants. The parameter θ_S is the constant risk premium of the stock, and θ_B is the risk premium of the bond fund.

Since B is a zero-coupon bond indexed on the interest rate r_t, the relation between their volatilities is given by:

$$\sigma_B = \frac{\sigma_r(1 - e^{-a_r D})}{a_r}.$$

Denote also by σ_{T-t} the volatility at time t of the zero-coupon bond maturing at time T:

$$\sigma_{T-t} = \frac{\sigma_r(1 - e^{-a_r(T-t)})}{a_r}, \tag{7.10}$$

which is decreasing with time (clearly σ_B coincides with σ_{T-t} when $T-t = D$).

Note that in this model, interest rates and stock index prices are negatively correlated[1]. Furthermore, the market is complete. Therefore, there exists a unique risk-neutral probability $\mathbb{Q}$ associated to two market risk premia, λ and λ_r, for which density η with respect to the initial probability P is given by:

$$\eta_t = \exp[-(\lambda W_t + \lambda_r W_{r,t}) - \frac{1}{2}(\lambda^2 + \lambda_r^2)t].$$

The premia λ and λ_r are determined from the relation:

$$\theta_S = \sigma_1 \lambda + \sigma_2 \lambda_r,$$
$$\theta_B = \sigma_B \lambda_r.$$

In this setting, B_T can be replicated using only the two assets C and B_D. These two assets span the bond market. As noted in [38], synthesizing B_T requires a positive weight on B_D and a weight on C, which is negative if $D < T$ and positive if $D > T$. This dynamic combination of fixed-income securities of different durations is referred to as the *passive immunization* (see Fong [239] or Fabozzi [215]).

[1] The instantaneous correlation between the interest rate and the stock index is equal to -$\sigma_r \sigma_2$. This assumption is not necessary to solve the optimization problems.

7.2.1.2 Optimal portfolios

We recall the standard results about optimal portfolio computation. Portfolio weights are denoted by x_C, x_S, and x_B. The portfolio value at time t is denoted by V_t. Therefore, the portfolio value V follows the dynamics:

$$\frac{dV_t}{V_t} = [r_t + x_S(t)\theta_S + x_B(t)\theta_B]dt + x_S(t)\sigma_1 dW + [x_S(t)\sigma_2 + x_B(t)\sigma_B]dW_r.$$

The investor's preferences are described by utility function U, which embeds his risk aversion. He has an initial capital denoted by V_0. He is assumed to maximize expected utility over the time horizon T. Therefore, his optimal portfolio weights are the solutions of the following problem:

$$\underset{x_C, x_S, x_B}{Max} \ \mathbb{E}\left[U\left(V_T\right)\right].$$

For different utility functions, we show below how the optimal dynamic factors depend on the investor risk aversion. In what follows, we present the solutions for four standard utility functions:

7.2.1.2.1 Logarithmic Utility

$$U(x) = \mathrm{Log}(x), x > 0, \tag{7.11}$$

so that the absolute risk aversion is $-U''(x)/U'(x) = 1/x$, and the relative risk aversion is constant and equal to $-xU''(x)/U'(x) = 1$. In that case, the optimal portfolio is called the numeraire portfolio (see Long [361]), or the growth-optimal portfolio (see Merton [389]). Its value at maturity T, denoted by $V_T^{\log}$, is given by:

$$V_T^{\log} = V_0 H_T,$$

where the numeraire portfolio H_T is

$$H_T = \left(\frac{\eta_T}{\exp\left(\int_0^T r_s ds\right)}\right)^{-1}. \tag{7.12}$$

The optimal weights at any time t are displayed in the next table.

In the growth-optimal portfolio, h_s and h_B represent the weights of the stock index and the constant maturation bond, respectively. In the case considered by [38], "this optimal portfolio is highly aggressive and thus levered (negative weight in cash). Increasing risk aversion implies decreasing weight in the growth optimal portfolio and consequently increasing weight in cash." Note that here the ratio x_B/x_S is constant since within the risky fund, the weights are the same regardless the level of risk aversion. Therefore, there

TABLE 7.1: Optimal weights for logarithmic utility function

x_C	$=$	$1-(x_S+x_B)$				
x_S	$=$	$h_S,$	with	h_S	$=\frac{\lambda}{\sigma_1}$	
x_B	$=$	$h_B,$	with	h_B	$=\frac{-\lambda\sigma_2+\lambda_r\sigma_1}{\sigma_1\sigma_B}$	

is no inconsistency with the mutual fund separation theorem. However, the ratio of *all* bonds B_T to stock increases with risk aversion.

Some properties of the numeraire portfolio:

Recall that the numeraire portfolio H_T is equal to $exp\left(\int_0^T r_s\, ds\right)/\eta_T$. Thus, since in particular $\left(\int_t^T r_s ds\right)$ has a Gaussian distribution, the ratio H_T^z/H_t^z is Lognormally distributed: it is equal to $exp[N_z(t,T)]$, where $N_z(t,T)$ has a Gaussian distribution.

Therefore, the expectation $\mathbb{E}\left[H_T^z/H_t^z\right]$ is defined by:

$$\mathbb{E}_t[\frac{H_T^z}{H_t^z}] = \exp[\mathbb{E}(N_z)(t,T) + \frac{1}{2}Var(N_z)(t,T)], \tag{7.13}$$

where

$$\mathbb{E}(N_z)(t,T) = z\ \Phi_{(t,T)} \text{ and } Var(N_z)(t,T) = z^2\Psi_{(t,T)}, \tag{7.14}$$

with

$$\Phi_{(t,T)} = (T-t)\ [b_r + (r_t - b_r)\frac{\sigma_{T-t}}{(T-t)\sigma_r} + \frac{\lambda^2+\lambda_r^2}{2}], \tag{7.15}$$

$$\Psi_{(t,T)} = (T-t)\times$$

$$\left[\frac{\sigma_r^2}{a_r^2}(1 - 2\frac{\sigma_{T-t}}{(T-t)\sigma_r} + \frac{\sigma_{2(T-t)}}{2(T-t)\sigma_r}) + (\lambda^2+\lambda_r^2) - 2\frac{\lambda_r\sigma_r}{a_r}(1 - \frac{\sigma_{T-t}}{(T-t)\sigma_r})\right]. \tag{7.16}$$

Denote respectively by $\varphi_{(t,T)}$ and $\psi_{(t,T)}$ the expectations $\mathbb{E}[H_t/H_T]$ and $\mathbb{E}[H_t \log(H_T/H_t)/H_T]$. We have:

$$\varphi_{(t,T)} = exp[-\Phi_{(t,T)} + \frac{1}{2}\Psi_{(t,T)}] \text{ and } \psi_{(t,T)} = \phi_{(t,T)} \times \left(\Phi_{(t,T)} - \Psi_{(t,T)}\right). \tag{7.17}$$

To simplify the notations, we denote respectively $\mathbb{E}(N_z)(T)$, $Var(N_z)(T)$, Φ_T, Ψ_T , φ_T, and ψ_T the values of $\mathbb{E}(N_z)(0,T)$, $Var(N_z)(0,T)$, $\Phi_{(0,T)}$, $\Psi_{(0,T)}$, $\varphi_{(0,T)}$, and $\psi_{(0,T)}$.

7.2.1.2.2 CRRA utility We now consider the utility function with a constant relative risk aversion, which generalizes the previous case.

$$U(x) = \frac{x^{1-\gamma}}{1-\gamma}, \gamma > 0, \tag{7.18}$$

so that $-xU''(x)/U'(x) = \gamma$. The value at maturity T, denoted by V_T^{CRRA} is

$$V_T^{CRRA} = (\lambda)^{-\frac{1}{\gamma}} H_T^{(\frac{1}{\gamma})}, \tag{7.19}$$

where λ is a Lagrange multiplier determined by the initial investment V_0. In this case, the optimal weights at any time t are deterministic functions of time (independent of the level of interest rate r). They are given in the next table.

TABLE 7.2: Optimal weights for CRRA utility function

x_C	$=$	$1 - \frac{1}{\gamma}(h_S + h_B) - (1 - \frac{1}{\gamma})(\frac{\sigma_{T-t}}{\sigma B})$
x_S	$=$	$\frac{1}{\gamma}(h_S)$
x_B	$=$	$(1 - \frac{1}{\gamma})(\frac{\sigma_{T-t}}{\sigma B}) + \frac{1}{\gamma}(h_B)$

Furthermore, the optimal bond over stock ratio is also time-dependent:

$$\frac{x_B(t)}{x_S(t)} = \frac{h_B}{h_S} + (\gamma - 1)\frac{1}{h_S}\frac{\sigma_{T-t}}{\sigma_B}.$$

Thus, whatever the time t and for any value of the parameters, this ratio is increasing in investor risk aversion. This property is consistent with current practice. Moreover, if γ is larger than 1, the weight in cash, x_C, is increasing in t, since it is a decreasing function of σ_{T-t} (see Equation (7.10)).

7.2.1.2.3 HARA utility The HARA utility function has a hyperbolic absolute risk aversion. It includes the CRRA utility function (described above), and the quadratic utility function as special cases. It is given by:

$$U(x) = \left(\frac{\gamma}{1-\gamma}\right)\left(\frac{x - x^*}{\gamma}\right)^{1-\gamma}, \tag{7.20}$$

where γ and x^* are two parameters that cannot both be negative. We also require that $x^* B_T(0) - V_0 < 0$. We have $-U''(x)/U'(x) = \gamma/(x - x^*)$ and $-xU''(x)/U'(x) = \gamma x/(x - x^*)$. Two cases arise according to the sign of γ. If $\gamma > 0$, the absolute risk aversion is decreasing in x. In this case, the

amount x^* represents a required lower bound for the terminal portfolio value ($x > x^*$). The initial investment V_0 is at least equal to the discounted value of this lower bound ($x^* B_T(0) < V_0$). If $\gamma < 0$, the absolute risk aversion is decreasing in x. The amount x^* is a required upper bound for the terminal portfolio value ($x < x^*$). The initial investment V_0 is at most equal to the discounted value of this lower bound ($x^* B_T(0) > V_0$).

In both cases, the risk aversion is an increasing function of the absolute value of γ. Note that the quadratic case is obtained with $\gamma = -1$ and $x^* > 0$. The optimal portfolio at maturity T, V_T^{HARA}, is given by:

$$V_T^{HARA} = (\mu)^{-\frac{1}{\gamma}} H_T^{(\frac{1}{\gamma})} + x^*.$$

This expression can be interpreted as follows: the optimal portfolio is a combination of a CRRA fund with γ parameter and a zero coupon bond yielding x^* at time T. The optimal portfolio weights for the HARA utility function are functions of $\varpi_t = 1 - B_{T-t}(t) x^* / V_t^{HARA}$.

TABLE 7.3: Optimal weights for HARA utility function

x_C	$=$	$1 - (x_S + x_B)$
x_S	$=$	$\varpi_t \times \frac{1}{\gamma}(h_S)$
x_B	$=$	$[\varpi_t \times (\frac{1}{\gamma})(h_B - \frac{\sigma_{T-t}}{\sigma_B})] + \frac{\sigma_{T-t}}{\sigma_B}$

In these expressions, $B_{T-t}(t)$ represents the value at time t of a zero-coupon bond maturing at time T. The factor ϖ_t represents the proportion at time t of the risky fund in the total portfolio value. Therefore, the ratio Bond over Stock is no longer deterministic. In this case, the optimal weights depend on market conditions (median, bullish, or bearish markets).

7.2.2 Exponential utility

Finally, we consider the exponential utility function:

$$U(x) = -\frac{e^{-ax}}{a},\ a > 0,$$

so that the absolute risk aversion is constant, $-U''(x)/U'(x) = a$. The optimal portfolio value for the exponential utility function, V_T^{exp}, is a function of the numeraire portfolio (see Equation (7.12)):

$$V_T^{exp} = A(V_0) + \frac{1}{a}\log(H_T),$$

where

$$A(V_0) = \frac{V_0 - \frac{1}{a}\mathbb{E}[\log(H_T)/H_T]}{E[1/H_T]}.$$

The computations of the expressions for the optimal weights are more involved:

- To compute the replicating strategies for a given optimal portfolio, first note that V_t/H_t is a martingale. Thus, we obtain the martingality relation
$$V_t/H_t = \mathbb{E}_{P,t}\left[V_T/H_T\right]. \tag{7.21}$$
- Therefore, it is necessary to compute conditional expectations of quantities which are functions of V_T/H_T. For this purpose, the conditional expectations of the numeraire portfolio have to be used.

Using the martingality relation, the value V_t^{exp} is given by:

$$V_t^{exp} = H_t\mathbb{E}_t\left[\frac{A(V_0) + \frac{1}{a}\log(H_T)}{H_T}\right].$$

Then, we obtain the expression of the optimal portfolio value at any time t of the investment period:

$$V_t^{exp} = \left[A(V_0)\varphi(t,r_t) + \frac{1}{a}\psi(t,r_t)\right] + \frac{1}{a}\ln(H_t)\varphi(t,r_t). \tag{7.22}$$

- To compute the optimal weights, we have to determine $\frac{dV_t^{\text{exp}}}{V_t^{\text{exp}}}$, and to search x_S, x_B, and x_C such that
$$\frac{dV_t^{exp}}{V_t^{exp}} = x_S\frac{dS_t}{S_t} + x_B\frac{dB_{D,t}}{B_{D,t}} + x_C\frac{dC_t}{C_t}.$$
- For this purpose, applying Ito's formula to Equation (7.22), we obtain:

$$\begin{aligned}
dV_t^{exp} = &\left[A(V_0)\frac{\partial\varphi}{\partial t}(t,r_t) + \frac{1}{a}\frac{\partial\psi}{\partial t}(t,r_t) + \frac{1}{a}\ln(H_t)\frac{\partial\varphi}{\partial t}(t,r_t)\right]dt \\
&+\frac{1}{2}\sigma_r^2[A(V_0)\frac{\partial^2\varphi}{\partial r^2}(t,r_t) + \frac{1}{a}\frac{\partial^2\varphi}{\partial r^2} + \frac{1}{a}\ln(H_t)\frac{\partial^2\varphi}{\partial r^2}]dt
\end{aligned}$$

$$\begin{aligned}
&+\left[A(V_0)\frac{\partial\varphi}{\partial r}(t,r_t) + \frac{1}{a}\frac{\partial\psi}{\partial r}(t,r_t) + \frac{1}{a}\ln(H_t)\frac{\partial\varphi}{\partial r}(t,r_t)\right]dr_t \\
&+\frac{1}{a}\varphi(t,r_t)\left[\frac{dH_t}{H_t} - \frac{1}{2H_t^2}d < H_t, H_t >\right],
\end{aligned}$$

where $< H_t, H_t >$ is the "instantaneous variance" of H_t. Define $\chi(t, r_t)$ by:

$$\chi(t,r_t) = A(V_0)\frac{\partial \varphi}{\partial r}(t,r_t) + \frac{1}{a}\frac{\partial \psi}{\partial r}(t,r_t) + \frac{1}{a}\ln(H_t)\frac{\partial \varphi}{\partial r}(t,r_t).$$

Then, dV_t^{exp} has the form:

$$dV_t^{exp} = V_t^{exp}\left[\frac{1}{a}\frac{\varphi(t,r_t)}{V_t^{exp}}(h_S\frac{dS_t}{S_t} + h_B\frac{dB_{D,t}}{B_{D,t}}) + \frac{\chi(t,r_t)}{V_t^{exp}}dr_t + (...)dt\right].$$

- Using this property, it is possible to determine the optimal weights by identifying the martingale components. This yields to the following relations:

$$x_S\sigma_1 = \frac{1}{a}\left(\frac{\varphi(t,r_t)}{V_t^{exp}}\right)h_S\sigma_1,$$

$$x_S\sigma_2 + x_B\sigma_B =$$

$$\frac{1}{a}\left(\frac{\varphi(t,r_t)}{V_t^{exp}}\right)h_S\sigma_2 + \frac{1}{a}\left(\frac{\varphi(t,r_t)}{V_t^{exp}}\right)h_B\sigma_B + \left(\frac{\chi(t,r_t)}{V_t^{exp}}\right)(-\sigma_r),$$

  from which we can compute the optimal weights for the exponential utility function.

- Finally, we deduce the optimal portfolio weights for the CARA utility function,

TABLE 7.4: Optimal weights for CARA utility function

x_C	$=$	$1-(x_S+x_B)$
x_S	$=$	$\frac{1}{a}\left(\frac{\varphi(t,T)}{V_{a,t}^{exp}}\right)h_S$
x_B	$=$	$\frac{1}{a}\left(\frac{\varphi(t,T)}{V_{a,t}^{exp}}\right)(\frac{\sigma_2-\sigma_1}{\sigma_B}h_S + h_B) - \left(\frac{\chi(t,T)}{V_{a,t}^{exp}}\frac{\sigma_r}{\sigma_B}\right)$

where

$$\chi_{(t,T)} = \left(\frac{1}{a}\right)\left[\varphi_{(t,T)}\left(\frac{\sigma_{T-t}}{\sigma_r}\right)\right]\left(1 + \Psi_{(t,T)} - \Phi_{(t,T)} - \frac{aV_0 - \psi_T}{\varphi_T} - \ln(H_t)\right).$$

Note that these optimal weights are deterministic functions of the instantaneous interest rate r_t, and the portfolio value $V_{a,t}^{exp}$. These functions depend only on market parameters. The ratio Bond over Stock is stochastic.

7.2.3 Sensitivity analysis

We consider a numerical base case. For simplicity, from now on we focus our attention on the CRRA utility function (the other cases can be treated in a similar way). For the base case, for the values of the short rate parameters, we use the estimates of Chan [114]: speed of convergence, $a_r = 12\%$, asymptotic short rate value, $b_r = 4\%$, and volatility, $\sigma_r = 4\%$ (see Equation (7.6)). The current instantaneous interest rate is $r_0 = 4\%$. The risk premia are $\theta_B = 1.5\%$, and $\theta_S = 6\%$; the index stock volatilities are $\sigma_1 = 19\%$, and $\sigma_2 = 6\%$ (see Equations (7.8) and (7.9)). Finally, the Bond fund constant duration is $D = 10$ years.

TABLE 7.5: Asset allocation sensitivities for CRRA utility

Market parameters and investor's typeγ	CASH	BONDS	STOCKS	$\frac{\text{Bonds}}{\text{Stocks}}$
Decrease of the volatility of Interest rate from 4 to 2	%	%	%	
3.52	-13	70	43	1.63
5.28	-8	80	28	2.86
10.57	-2	88	14	6.28
Decrease of the speed of convergence a_r from 12 to 10				
3.52	-10	65	45	1.44
5.28	-5	75	30	2.50
10.57	0	85	15	5.66
Change in the value of $b_r$from 4 to 2				
3.52	-10.5	65.5	45	1.45
5.28	-6	76	30	2.53
10.57	-2	87	15	5.80

A decreasing of the volatility of the interest rate shifts money from stock to cash and bonds (the instantaneous bond return is less risky). Decreasing the speed of convergence a_r is similar to increasing interest rate volatility, since interest rates converge slowly to their long run value b_r. Decreasing the value of b_r has almost no impact for the CRRA case, since portfolio weights are independent of the interest rate level.

7.2.3.1 Utility of the optimum portfolio

Consider an investor with a HARA utility function, with relative risk aversion γ. His time horizon is T. In this case, the discounted expected utility (computed at time $t = 0$) is given by:

$$\mathbb{E}[U_\gamma(V^*_{(\gamma,T)}; V_0)] = \left(\frac{\gamma^\gamma}{1-\gamma}\right)[(V_0 - x^* \alpha_T)]^{(1-\gamma)} \tag{7.23}$$

$$\exp\left[\gamma\left(\mathbb{E}(N_z)(T) + \frac{1}{2}Var(N_z)(T)\right)\right], \tag{7.24}$$

and, for the special case CRRA, we obtain:

$$\mathbb{E}[U_\gamma(V^*_{(\gamma,T)}; V_0)] = \left(\frac{\gamma^\gamma}{1-\gamma}\right) V_0^{1-\gamma} \exp\left[\gamma\left(\mathbb{E}(N_z)(T) + \frac{1}{2}Var(N_z)(T)\right)\right], \tag{7.25}$$

with $z = (1-\gamma)/\gamma$, where V_0 is the initial investment, and where $E(N_z)(T)$ and $Var(N_z)(T)$ are defined by relation (7.14).

Consider now an investor with a CARA utility function, with absolute risk aversion a. We have:

$$\mathbb{E}[\widehat{U}_a(V^*_{(a,T)}; V_0)] = -\frac{1}{a}\exp\left[\frac{-aV_0 + \beta_T}{\alpha_T}\right] \times \alpha_T. \tag{7.26}$$

Under the same numerical assumptions of the previous numerical base case, consider a financial institution that offers to its clients three standardized portfolios: the first one is aggressive (45% Stock), the second one is a moderate portfolio (30% Stock), and the third is a conservative portfolio (15% Stock). Assuming CRRA utility functions, it is possible to recover two unknowns: the values of the risk aversion and of the investment period corresponding to this portfolio. For this example, we find $T = 25$ years, and we provide below values of the risk aversion which best fit these three portfolios.

TABLE 7.6: Asset allocations for agressive, moderate and conservative investors (CRRA utility function)

CASH %	BONDS %	STOCKS %	Ratio of Bonds to Stocks		Investor type γ
-10	65	45	1.44	Aggressive	3.52
-6	76	30	2.53	Moderate	5.28
-2	87	15	5.8	Conservative	10.57

7.2.4 Distribution of the optimal portfolio return

The performance of an optimal strategy can be illustrated by the distribution of return at maturity. Below, this distribution is computed for the CRRA utility function (the other cases can be derived in a similar way).

From Equation (7.19), we deduce the optimal portfolio return V_T^{CRRA}/V_0:

$$R(\gamma) = H_T^{(1/\gamma)}/\mathbb{E}[H_T^{(1/\gamma-1)}].$$

Since $H_T^{(1/\gamma)}$ has a lognormal distribution, the cumulative distribution function $F_{R(\gamma)}(x)$ of the return $R(\gamma)$ can be computed explicitly. Let Y_T denote the random variable such that $H_T = \exp[Y_T]$. The distribution of Y_T is Gaussian. Denote respectively by m_T and s_T its expectation and standard deviation.

The expectation $\mathbb{E}\left[H_T^{(1/\gamma-1)}\right]$ is defined by:

$$\mathbb{E}\left[H_T^{(1/\gamma-1)}\right] = \exp\left[\mathbb{E}(N_{(1-\gamma)/\gamma}) + \frac{1}{2}Var(N_{(1-\gamma)/\gamma})\right],$$

with

$$\mathbb{E}(N_z) = z\ \Phi_T \text{ and } Var(N_z) = z^2\Psi_T. \tag{7.27}$$

Then, we have:

$$F_{R(\gamma)}\ (x) = \mathbb{P}\left[\frac{H_T^{(1/\gamma)}}{\mathbb{E}\left[H_T^{(1/\gamma-1)}\right]} \leq x\right] = \mathbb{P}\left[X \leq \frac{\gamma \text{Log}\left[x\ \mathrm{E}\left[H_T^{(1/\gamma-1)}\right]\right] - m_T}{s_T}\right],$$

where X is a random variable with a standard Gaussian distribution. Note that m_T and s_T are constant, which only depend on market parameters (but not on risk aversion γ).

The inverse cumulative distribution functions are displayed in the next figure in the base case for three values of γ ($\gamma = 3.52$, $\gamma = 5.28$, and $\gamma = 10.57$) and for a time horizon $T = 20$ years.

At a given probability level 0.98, an investor with a aggressive portfolio has a guarantee to recover only 70% of his initial investment. However, this investor has a 20% chance to multiply his initial investment by a factor 5. If this investor selects a conservative portfolio, he has a guarantee (up to 98%) to increase by half the value of his initial portfolio value. However, in this case, the probability to multiply his initial investment by 5 is almost zero (the probability to multiply the initial investment by 3 is only about 10%). Finally, the investor selecting the moderate portfolio (intermediary curve) has a 98% probability to recover his initial investment after 20 years.

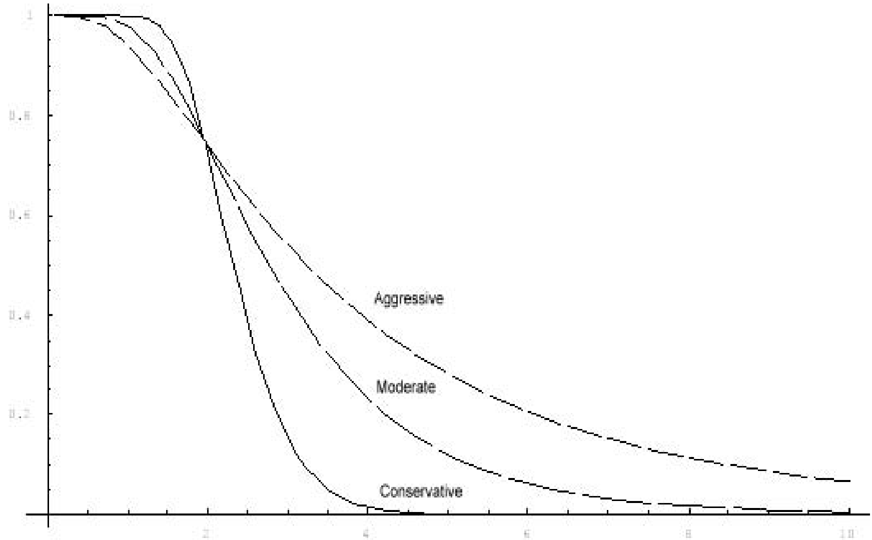


FIGURE 7.3: Inverse cumulative distribution of the return at maturity

7.3 Further reading

Long-term management depends on a large variety of factors. These include: income, saving capacity, age, gender, know-how, experience, liquidity, real estate, investment horizon, attitudinal and personality factors, investment objectives (e.g. planned projects or retirements), initial investable assets and additional funds. Campbell and Viceira [103] discuss to what extent life-cycle portfolio choice and saving are affected by the variables mentioned above.

The influence of risk aversion has been examined by Kallberg and Ziemba [315]; the optimal portfolio is more sensitive to the value of risk aversion than to the functional form of utility function.

Similarly, Brennan and Xia [89] highlight the importance of considering investors' time horizons in the analysis of optimal portfolio policies.

Long-term portfolio optimization with fixed incomes is studied in Brennan and Xia [89], who examine the Bond-Stock Mix when stochastic interest rates are involved. Fabozzi [215] provides an overview on bond markets analysis. Sorensen [473] and Lioui and Poncet [357] analyze optimal portfolio choice and fixed income management under stochastic interest rates.

Battocchio *et al.* [50] examine in particular the role of the decumulation

phase in the determination of the optimality of asset allocation under mortality risk.

Note also that financial institutions typically offer a limited number of standardized portfolios which imperfectly match investor preferences. Jensen and Sorensen [302] and De Palma and Prigent [161] quantify the efficiency losses of an investor acquiring a standardized (e.g., conservative, balanced, and aggressive) versus a customized portfolio. A standardized portfolio may differ from the optimal one by its risk exposure or its time horizon. Numerical results show that the monetary losses from not having access to a customized portfolio can be substantial, in particular when the time horizon of the investor differs from the one selected by the financial institution.

Chapter 8

Optimization within specific markets

As seen in Chapter 6, the investor's utility maximization of his terminal wealth can be solved in the continuous-time setting. Using methods of stochastic optimal control, Merton ([385], [386]) proves that the value function is the solution of a non-linear partial differential equation: the Bellman equation. Then, closed-form solutions are available for the HARA utility. However, the dynamic programming method is based on Markovian assumptions. To avoid this hypothesis, another approach has been introduced: the duality portfolio characterization by using the martingale measures, the so-called risk-neutral measures. For complete markets, this set is reduced to one point and the optimal solution is determined from the fundamental result: the terminal wealth of the optimal portfolio is equal (up to a multiplicative constant) to the marginal utility inverse of the density of the martingale measure. This method, illustrated in Chapter 6, has been introduced by Pliska [410], Cox and Huang ([132],[133]), and Karatzas *et al.* [319]. This is in line with the optimal investment problem for a one-period model with a finite set of random events solved by introducing the Arrow-Debreu state prices.

However:

- The assumption of financial market completeness is very strong. Roughly speaking, it is supposed that there are as many assets as random sources.
- In addition, several market "frictions" must also be taken into account:

  - Constraints, such as no-short selling;

  - Transaction costs;

  - Labor income stream;

  - Partial information about security prices.

Thus, specific methods must be introduced to examine such optimization problems.

8.1 Optimization in incomplete markets

He and Pearson ([288], [289]) have studied this problem both in discrete-time and continuous-time frameworks. Karatzas et al. [319] have shown how expected utility maximization can be solved by martingale methods and convex duality. In what follows, first a general result, due to Kramkov and Schachermayer [335], is presented. Then, some standard financial models are considered and explicit optimal solutions are detailed.

8.1.1 General result based on martingale method

- *Price process:* Consider a financial market with $(d+1)$ securities, one riskless bond with constant rate r, and d stocks with price process $(S_{i,t})_{i,t}$, $t \in [0,T]$, and $1 \leq i \leq d$. Note that the time horizon can be also infinite. The process S is assumed to be a semimartingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$.

- *Portfolio strategy and value:* Recall that a self-financing portfolio strategy $(\theta_{i,t})_{i,t}$, where $\theta_{i,t}$ denotes the amount invested on asset i at time t, is a predictable process, which is integrable w.r.t. the price process S. The portfolio value process $(V_t)_t$, associated to strategy $(\theta_{i,t})_{i,t}$, is given by: for $t \in [0,T]$,

$$V_t = V_0 + \sum_{i=1}^{d} \int_0^t \theta_{i,s} dS_{i,s}. \tag{8.1}$$

Denote by $\mathcal{V}(V_0)$ the family of wealth processes $(V_t)_t$ such that for $t \in [0,T]$, $V_t \geq 0$ and with initial value V_0.

- *Equivalent local martingale:* A probability measure $\mathbb{Q}$, equivalent to $\mathbb{P}$, is called an equivalent local martingale measure if any process V in $\mathcal{V}(1)$ is a local martingale w.r.t. $\mathbb{Q}$. Note that when the process S is bounded (resp. locally bounded), then under an equivalent martingale measure, the process S is a martingale (resp. locally martingale) (see Delbaen and Schachermayer [155] for more details about such property). Denote by $\mathcal{M} = \mathcal{M}_e(S)$ the set of equivalent local martingale measures which is assumed to be non-empty due to the absence of arbitrage opportunities, as shown by Harrison *et al.* ([284], [285]).

- *Expected utility maximization:* The investor has a utility function on wealth $U : (0, +\infty) \to \mathbb{R}$. The investor searchs for the value function associated to the primal problem:

$$\mathcal{J}(V_0) = \sup_{V_T \in \mathcal{V}(V_0)} \mathbb{E}\left[U(V_T)\right]. \tag{8.2}$$

Assumptions (A) on utility function: The function U is defined on $\mathbb{R}^+$, strictly increasing, strictly concave, continuously differentiable, and such that:

$$U'(0 = \lim_{x \to 0^+} U'(x) = +\infty,$$
$$U'(\infty) = \lim_{x \to \infty} U'(x) = 0.$$

The value function $\mathcal{J}$ is supposed to satisfy for some $x > 0$, $\mathcal{J}(x) < \infty$. In order to solve the optimization problem in this general framework, Kramkov and Schachermayer [335] introduce the following key assumption:

- *Asymptotic elasticity of the utility function:* The utility function U has asymptotic elasticity AE(U) strictly smaller than 1 if:

$$AE(U) = \lim\sup_{x \to \infty} \frac{xU'(x)}{U(x)} < 1. \tag{8.3}$$

Examples of this utility function are the logarithm and power utilities: for $U(x) = \ln x$, $AE(U) = 0$, and for $U(x) = \frac{x^\alpha}{\alpha}$ with $0 < \alpha < 1$, $AE(U) = \alpha$. However, for instance, the utility function $U(x) = \frac{x}{\ln x}$ has AE(U) equal to 1.

Kramkov and Schachermayer [335] show that the condition $AE(U) < 1$ is necessary and sufficient to get the following properties under the previous assumption on price process S:

- 1) The value function $\mathcal{J}$ is a utility function which is increasing, strictly concave, continuously differentiable, and such that:

$$\lim_{x \to 0^+} \mathcal{J}'(x) = +\infty \text{ and } \lim_{x \to \infty} \mathcal{J}'(x) = 0.$$

- 2) The optimization problem has a solution.

Note also that, either the utility function $U(.)$ satisfies $AE(U) < 1$, which implies that $AE(\mathcal{J}) < 1$, or $AE(U) = 1$ in which case there exists an $\mathbb{R}$-valued price process S which is continuous, induces a complete market, and is such that $\mathcal{J}$ is not strictly concave (more precisely, there exists x_0 such that $\mathcal{J}(x)$ is a straightline with slope one for $x > x_0$).

- *Legendre-transform and conjugate utility function:* As seen in Chapter 6, it is useful to introduce the conjugate function of the utility function U:

$$\widehat{U}(y) = \max_{x>0} \left[U(x) - xy\right], \ y > 0. \tag{8.4}$$

The function $\widehat{U}$ is the Legendre-transform of the function $-U(-x)$ (see Rockafellar [426]). If the utility function U satisfies assumption (A), the function $\widehat{U}$ is continuously differentiable, decreasing, strictly convex and such that:

$$\lim_{x \to 0^+} \widehat{U}(x) = \lim_{x \to \infty} U(x) \text{ and } \lim_{x \to \infty} \widehat{U}(x) = \lim_{x \to 0^+} U(x), \tag{8.5}$$

$$\lim_{x \to 0^+} \widehat{U}'(x) = -\infty \text{ and } \lim_{x \to \infty} \widehat{U}'(x) = 0. \tag{8.6}$$

Furthermore, we have the following bidual property:

$$U(x) = \max_{y>0} \left[\widehat{U}(y) + xy\right], \ x > 0. \tag{8.7}$$

The derivative of $U(.)$ is the inverse function of the negative of the derivative of $\widehat{U}(.)$. Denote it by J. We then have: $J = -\widehat{U}' = (U')^{-1}$.

Example 8.1
Consider the power utility $U(x) = \frac{x^\alpha}{\alpha}$ with $0 < \alpha < 1$. Then,

$$\widehat{U}(y) = \frac{1-\alpha}{\alpha} y^{\frac{\alpha}{1-\alpha}}.$$

Consider the exponential utility $U(x) = -\frac{e^{-ax}}{a}$ with $0 < a$. Then,

$$\widehat{U}(y) = \frac{y}{a}\left(\ln y - 1\right).$$

□

- *Optimal solutions:* We refer to [335] for the three following theorems and their detailed proofs, corresponding first to the complete case, and second to the incomplete case with $AE(U) < 1$ or $AE(U) = 1$.

THEOREM 8.1 Complete case

Suppose that previous assumptions are satisfied, and also that $\mathcal{M} = \{\mathbb{Q}\}$. Denote:

$$\widehat{\mathcal{J}}(y) = \mathbb{E}\left(\widehat{U}\left[y\frac{d\mathbb{Q}}{d\mathbb{P}}\right]\right). \tag{8.8}$$

We have:
(i) The functions $\mathcal{J}$ and $\widehat{\mathcal{J}}$ are finite:

$$\mathcal{J}(x) < \infty, \forall x > 0 \text{ and } \widehat{\mathcal{J}}(y) < \infty, \forall y > 0 \text{ sufficiently large.}$$

Denote $y_0 = \inf\left\{y : \widehat{\mathcal{J}}(y) < \infty\right\}$. The function $\widehat{\mathcal{J}}(y)$ is continuously differentiable and strictly convex on $]y_0, \infty[$. Denote $x_0 = \lim_{y\to y_0}\left(-\widehat{\mathcal{J}}'(y)\right)$. The function $\mathcal{J}$ is continuously differentiable on $]0, \infty[$ and strictly concave on $]0, x_0[$. The value functions $\mathcal{J}$ and $\widehat{\mathcal{J}}$ are conjugate:

$$\widehat{\mathcal{J}}(y) = \max_{x>0}\left[\mathcal{J}(x) - xy\right], \ y > 0, \tag{8.9}$$

$$\mathcal{J}(x) = \max_{y>0}\left[\widehat{\mathcal{J}}(y) + xy\right], \ x > 0, \tag{8.10}$$

and their derivatives satisfy:

$$\lim_{x\to 0^+} \mathcal{J}'(x) = \infty \text{ and } \lim_{y\to\infty} \widehat{\mathcal{J}}'(y) = 0.$$

(ii) If $x < x_0$, the optimal solution $V_T^*(x)$ is given by:

$$V_T^*(x) = I\left(y\frac{d\mathbb{Q}}{d\mathbb{P}}\right), \text{ for } y < y_0,$$

where x and y satisfy $y = \mathcal{J}'(x)$, or equivalently $x = -\widehat{\mathcal{J}}'(y)$. Note that, in that case, the optimal solution process $V^*(x)$ is a uniformly integrable martingale under the risk-neutral probability $\mathbb{Q}$.

(iii) For $0 < x < x_0$ and for $y > y_0$, we have:

$$\mathcal{J}'(x) = \mathbb{E}\left[\frac{V_T^*(x)U'\left(V_T^*(x)\right)}{x}\right], \ \widehat{\mathcal{J}}'(y) = \mathbb{E}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\widehat{U}'\left(y\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right].$$

In order to study the incomplete case, we introduce:

- The family $\mathcal{Y}(y)$ of non-negative semimartingales with $Y_0 = y$ and such that, for any $X \in \mathcal{V}(1)$, the product XY is a supermartingale. Note that this set contains the density processes of all equivalent local martingales $\mathbb{Q}$.

- The value function of the dual optimization problem is denoted and defined by:

$$\widehat{\mathcal{J}}(y) = \inf_{Y\in\mathcal{Y}(y)} \mathbb{E}\left(\widehat{U}\left[Y_T\right]\right). \tag{8.11}$$

  For the complete case, we have also $\widehat{\mathcal{J}}(y) = \mathbb{E}\left(\widehat{U}\left[y\frac{d\mathbb{Q}}{d\mathbb{P}}\right]\right)$. For the incomplete case, they may differ.

Since the functions $\mathcal{J}$ and $-\widehat{\mathcal{J}}$ are concave, the right-continuous versions of their derivatives $\mathcal{J}'$and $-\widehat{\mathcal{J}}'$ exist. Recall that the asymptotic elasticity $AE(\mathcal{J})$ is defined by:

$$AE(\mathcal{J}) = \lim\ \sup_{x\to\infty} \frac{x\mathcal{J}(x)}{\mathcal{J}(x)}.$$

The following two theorems (see [335]) provide general results for the incomplete case.

THEOREM 8.2 Incomplete case with general utility function U

Under previous assumptions, we have:

(i) The function $\mathcal{J}$ *has finite values,* $\mathcal{J}(x) < \infty, \forall x > 0$*, and the function* $\widehat{\mathcal{J}}$ *is finite* $\forall y > y_0$*,* y_0 *sufficiently large. Denote* $y_0 = \inf\left\{y : \widehat{\mathcal{J}}(y) < \infty\right\}$*. The function* $\mathcal{J}$ *is continuously differentiable on* $]0, \infty[$*, and the function* $\widehat{\mathcal{J}}(y)$ *is strictly convex on* $]0, \infty[$*. The value functions* $\mathcal{J}$ *and* $\widehat{\mathcal{J}}$ *are conjugate:*

$$\widehat{\mathcal{J}}(y) = \max_{x>0} \left[\mathcal{J}(x) - xy\right], \; y > 0, \tag{8.12}$$

$$\mathcal{J}(x) = \max_{y>0} \left[\widehat{\mathcal{J}}(x) + xy\right], \; x > 0. \tag{8.13}$$

and their derivatives satisfy:

$$\lim_{x \to 0^+} \mathcal{J}'(x) = \infty \text{ and } \lim_{y \to \infty} \widehat{\mathcal{J}}'(y) = 0, \tag{8.14}$$

(ii) If $\widehat{\mathcal{J}}(y) < \infty$*, then the optimal solution* $V_T^*(x)$ *exists and is unique.*

THEOREM 8.3 Incomplete case with $AE(U) < 1$

Under previous assumptions and the additional hypothesis $AE(U) < 1$*, we have:*

i) The function $\widehat{\mathcal{J}}$ *is finite:* $\widehat{\mathcal{J}}(y) < \infty, \forall y > 0$ *sufficiently large. The functions* $\mathcal{J}$ *and* $\widehat{\mathcal{J}}$ *are continuously differentiable on* $]0, \infty[$*. The functions* $\mathcal{J}'$ *and -*$\widehat{\mathcal{J}}'$ *are strictly decreasing and satisfy*

$$\lim_{x \to \infty} \mathcal{J}'(x) = 0 \text{ and } \lim_{y \to 0^+} -\widehat{\mathcal{J}}'(y) = +\infty. \tag{8.15}$$

The asymptotic elasticity $AE(\mathcal{J})$ *is also smaller than* 1 *and: (notation* $x^+ = max(x, 0)$*)*

$$AE(\mathcal{J})^+ \leq AE(U)^+ < 1. \tag{8.16}$$

(ii) The optimal solution V_T^* *exists and is unique. If* $Y^*(y) \in \mathcal{Y}(y)$ *is the optimal solution of problem (8.11), where* $y = \mathcal{J}'(x)$*, we have the dual relation:*

$$V_T^*(x) = J\left(Y_T^*(y)\right). \tag{8.17}$$

The process $V_T^*(x)Y_T^*(y)$ *is a uniformly integrable martingale on* $[0, T]$*.*

(iii) The relations between $\mathcal{J}$*,* $\widehat{\mathcal{J}}$ *and* V_T^**,* Y_T^* *are:*

$$\mathcal{J}(x) = \mathbb{E}\left[\frac{V_T^*(x)U'\left(V_T^*(x)\right)}{x}\right]; \; \widehat{\mathcal{J}}'(y) = \mathbb{E}\left[\frac{Y_T^*(y)\widehat{U}'\left(Y_T^*(y)\right)}{y}\right]. \tag{8.18}$$

(iv) The value function $\widehat{\mathcal{J}}$ is given by:

$$\widehat{\mathcal{J}}(y) = \inf_{\mathbb{Q}\in\mathcal{M}} \mathbb{E}\left[\mathcal{J}\left(y\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right]. \tag{8.19}$$

Example 8.2 **One-period model**

Consider a one-period model where the investor has a logarithmic utility function $U(x) = \ln x$. Thus, we have:

$$AE(U) = \lim \sup_{x\to+\infty} \frac{x^{\frac{1}{x}}}{\ln x} = 0 < 1 \text{ and } \begin{cases} J(y) = \frac{1}{y}, \\ \widehat{U}(y) = -\ln y - 1. \end{cases}$$

Suppose that the financial market contains a riskless asset, taken as numeraire, and a stock S with only three return values. Set $S_0 = 1$. The stock value at maturity T is given by:

$$S_T = \begin{cases} u \text{ ("up") with probability } p, \\ m \text{ ("mean") with probability } 1-p-q, \\ d \text{ ("down") with probability } q. \end{cases}$$

with $d < 1, m < u$. Assume that $m = 1$.

Then, a supermartingale Y in $\mathcal{Y}(y)$ has the following form: $Y_0 = y$ and

$$Y_T = \begin{cases} y_u \text{ ("up") with probability } p, \\ y_m \text{ ("mean") with probability } 1-p-q, \\ y_d \text{ ("down") with probability } q, \end{cases}$$

where y_d, y_m, and y_u are non-negative scalars.

The primal optimization problem is associated to the value function $\mathcal{J}$:

$$\mathcal{J}(x) = \sup_{V_T\in\mathcal{V}(x)} \mathbb{E}\left[\ln V_T\right].$$

The dual problem is associated to the value function $\widehat{\mathcal{J}}$:

$$\widehat{\mathcal{J}}(y) = \inf_{Y\in\mathcal{Y}(y))} \mathbb{E}\left[-\ln V_T - 1\right].$$

Since Y and $Y.S$ are in $\mathcal{Y}(y)$, $\widehat{\mathcal{J}}(y)$ can be written as:

$$\begin{aligned}\widehat{\mathcal{J}}(y) &= -1 - \sup_{y_d,y_m,y_u} \mathbb{E}\left[-\ln Y_T\right], \\ &= -1 - \sup_{y_d,y_m,y_u} \left[q\ln y_d + (1-q-p)\ln y_m + p\ln y_u\right],\end{aligned}$$

with

$$\begin{cases} \mathbb{E}\left[Y_T\right] = qy_d + (1-q-p)y_m + py_u \leq y, \\ \mathbb{E}\left[Y_T S_T\right] = qdy_d + (1-q-p)y_m + puy_u \leq y, \\ y_d, y_m, y_u \geq 0. \end{cases}$$

Note that the optimum (y_d^*, y_m^*, y_u^*) is necessarily such that the previous constraints are true equalities. Then, we deduce:

$$Y_T^* = \begin{cases} y_u^* = y\frac{(p+q)(1-d)}{p(u-d)} \\ y_m^* = y \\ y_d^* = y\frac{(p+q)(u-1)}{q(u-d)} \end{cases} = y\begin{cases} \frac{(p+q)(1-d)}{p(u-d)} \\ 1 \\ \frac{(p+q)(u-1)}{q(u-d)} \end{cases}.$$

Finally, since $V_T^*(x) = J\left(Y_T^*(y)\right)$ where $y = \mathcal{J}'(x) = \frac{1}{x}$, we deduce:

$$V_T^* = x\begin{cases} \frac{p(u-d)}{(p+q)(1-d)} \\ 1 \\ \frac{q(u-d)}{(p+q)(u-1)} \end{cases}.$$

□

Another standard example is based on a $d-$dimensional Brownian motion with dimension d higher than the number of available assets.

Example 8.3 **Multidimensional Brownian motion**

Consider now a continuous-time model where the investor still has a logarithmic utility function $U(x) = \ln x$. The financial market still contains a riskless asset, taken as numeraire, and one stock S. Set $S_0 = 1$. Assume that S is solution of the SDE:

$$dS_t = S_{t_-}.\left(\mu dt + \sigma_1 dW_{1,t} + \sigma_2 dW_{2,t}\right), \tag{8.20}$$

where $W = (W_1, W_2)$ is a $2-$dimensional standard Brownian motion. Thus, the price S is equal to:

$$S_t = \exp\left(\left[\mu - \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right)\right]t + \sigma_1 W_{1,t} + \sigma_2 W_{2,t}\right), \tag{8.21}$$

where μ, σ_1, and σ_2 are constant with $\sigma_1 > 0$ and $\sigma_2 > 0$.

Due to the predictable representation theorem, the process Y has the following form:

$$Y_t = Y_0 \exp\left(\int_0^t \left[\mu^Y(s) - \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right)\right]ds + \int_0^t \sigma_{1,s}^Y dW_{1,s} + \int_0^t \sigma_{2,s}^Y dW_{2,s}\right). \tag{8.22}$$

The processes Y and $Y.S$ are in $\mathcal{Y}(y)$. Therefore, for any T,

$$\mathbb{E}\left[Y_T \,|Y_0 = y\right] = y\exp\left(\int_0^T \mu^Y(t)dt\right) \leq y,$$

$$\mathbb{E}\left[Y_T S_T \,|Y_0 = y\right] = y\exp\left(\int_0^T \left[\mu + \mu^Y(t) + \sigma_1\sigma_{1,t}^Y + \sigma_2\sigma_{2,t}^Y\right]dt\right) \leq y.$$

Thus $\mu^Y(.) \equiv 0$. Then, we have:

$$\mathbb{E}(\ln V_T) = \ln y - \frac{1}{2}\int_0^T \left[\left(\sigma_{1,t}^Y\right)^2 + \left(\sigma_{2,t}^Y\right)^2\right] dt.$$

Consequently,

$$\mu + \sigma_1 \sigma_{1,t}^Y + \sigma_2 \sigma_{2,t}^Y = 0.$$

Then,

$$\mathbb{E}(\ln V_T) = \ln y - \frac{1}{2}\int_0^T \left[\mu^2 + 2\mu\sigma_2\left(\sigma_{2,t}^Y\right) + \left(\sigma_{2,t}^Y\right)^2\left(\sigma_1^2 + \sigma_2^2\right)\right] / \left(\sigma_1\right)^2 dt.$$

Finally, we deduce that the optimal solution is given by:

$$\begin{aligned}
\sigma_{1,t}^Y &= -\frac{\mu\sigma_2}{(\sigma_1^2 + \sigma_2^2)};\ \sigma_{2,t}^Y = -\frac{\mu\sigma_1}{(\sigma_1^2 + \sigma_2^2)},\\
Y_T^* &= y \exp\left[\frac{\mu}{(\sigma_1^2 + \sigma_2^2)}\left(-\frac{T}{2}\mu - \sigma_1 W_{1,T} - \sigma_2 W_{2,T}\right)\right],\\
\widehat{\mathcal{J}}(y) &= -1 - \mathbb{E}[\ln Y_T^*] = -1 - \ln y + \frac{\mu^2}{(\sigma_1^2 + \sigma_2^2)}.
\end{aligned}$$

Since $y = \mathcal{J}'(x) = \frac{1}{x}$, we deduce

$$V_T^* = V_0 \exp\left[\frac{\mu}{(\sigma_1^2 + \sigma_2^2)}\left(\frac{T}{2}\mu + \sigma_1 W_{1,T}\, \sigma_2 W_{2,T}\right)\right].$$

□

REMARK 8.1 The optimal portfolio value V_T^* is an increasing function of the stock price at maturity. Indeed, using the relation:

$$S_T \exp\left(-\left[\mu - \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right)\right] T\right) = \exp\left(\sigma_1 W_{1,t} + \sigma_2 W_{2,T}\right),$$

we deduce:

$$V_T^* = V_0 \times v_T \times S_T^{\frac{\mu}{(\sigma_1^2 + \sigma_2^2)}},$$

where v_T is a deterministic function given by:

$$v_T = \exp\left[\frac{\mu^2 T}{2\left(\sigma_1^2 + \sigma_2^2\right)}\right] \exp\left[\frac{-\mu}{(\sigma_1^2 + \sigma_2^2)}\left(\mu - \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2\right) T\right)\right].$$

The concavity/convexity of V_T^* only depends on the comparison between μ and $\sigma_1^2 + \sigma_2^2$.

□

8.1.2 Dynamic programming and viscosity solutions

The standard dynamic approach can be extended to portfolio analysis within incomplete markets. The value function can be characterized as a viscosity solution of the associated Bellman equation. In general, the value function is not smooth. However, as shown by Pham [408], for CRRA utility functions, the non-linearity of the Bellman equation can be reduced by using a logarithmic transformation, as introduced by Fleming [230] in a stochastic control setting. This allows us to get a semilinear equation with quadratic growth on the derivative term.

8.1.2.1 Stochastic volatility

In what follows, the results of Pham [408] are presented (see also Zariphopoulou [510] for the univariate case).

- *The financial market* contains a riskless asset with rate r. The price S of the d risky assets is assumed to be a semimartingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$, and to be the solution of the SDE:

$$dS_t = diag(S_t).\left[(\mu\left(F_t\right)dt + \sigma_1\left(F_t\right)dW_{1,t} + \sigma_2\left(F_t\right)dW_{2,t}\,)\right], \tag{8.23}$$

where W_1 is a $d-$dimensional standard Brownian motion, and W_2 is an $m-$dimensional standard Brownian motion, independent of W_1.

$Diag(S_t)$ is the diagonal matrix $d \times d$ matrix with $m_{i,i} = S_{i,t}$.

The function $\sigma_1\left(.\right)$ is $\mathbb{R}^d$-valued, and $\sigma_2\left(.\right)$ is a $d \times m$ matrix. Both $\sigma_1\left(.\right)$ and $\sigma_2\left(.\right)$ are assumed to be continuous functions.

The process F refers to stochastic factors and is defined from:

$$dF_t = \eta\left(F_t\right)dt + dW_{1,t}\,,$$

where the function $\eta\left(.\right)$ is assumed to be Lipschitzian.

Denote $\widetilde{\mu}(y) = \mu(y) - r\mathbb{I}$, the excess rate of return w.r.t. the riskless rate ($\mathbb{I}$ is the vector of one in $\mathbb{R}^d$).

Denote also $\Sigma(y) = [\sigma_1\left(y\right)\sigma_2\left(y\right)]$ the matrix-valued volatility of the risky assets. The matrix Σ is of full rank equal to d. Denote:

$$\beta(y) = \inf_{w\in\mathbb{R}^d} \frac{\left|\left|{}^t\Sigma(y)w\right|\right|^2}{\left|\left|w\right|\right|^2}.$$

Assume that there exists a positive constant C such that:

$$\begin{aligned}
\left|\left|\widetilde{\mu}(y)\right|\right|/\sqrt{\beta(y)} &\leq C(1+\left|\left|y\right|\right|), \\
\left|\left|\sigma(y)\right|\right|/\sqrt{\beta(y)} &\leq C(1+\left|\left|y\right|\right|).
\end{aligned}$$

- *The investor* has a weighting process $(w_t)_t$ which is, as usual, assumed to be predictable and integrable:

$$\sup_{t\in[0,T]} \mathbb{E}\left[\exp\left(a\left|\left|\Sigma(y)^t w\right|\right|\right)\right] < \infty, \text{ for some constant } a > 0.$$

Denote by $\mathcal{A}$ the set of admissible strategies. The investor has a power utility $U(x) = \frac{x^\alpha}{\alpha}$, with $0 < \alpha < 1$.

The value function of the investor is defined by: for any $(t, x, f) \in [0, T] \times \mathbb{R}^+ \times \mathbb{R}^d$,

$$\mathcal{J}(t, x, f) = \sup_{w\in\mathcal{A}} \mathbb{E}\left[U(V_T)\,|V_t = x, F_t = f\right]. \tag{8.24}$$

- *The dynamic programming equation* (the Hamilton-Jacobi-Bellman equation) associated to the stochastic control problem is the non linear PDE:

$$\frac{\partial \mathcal{J}}{\partial t} + rx\frac{\partial \mathcal{J}}{\partial x} +^t \eta(f) D_f \mathcal{J} + \frac{1}{2}\Delta_f \mathcal{J} + \tag{8.25}$$
$$\max\left[{}^t w\mu(f)x\frac{\partial \mathcal{J}}{\partial x} + \frac{1}{2}\left|\left|\Sigma(f)^t w\right|\right|^2 x^2 \frac{\partial^2 \mathcal{J}}{\partial t^2} + \ {}^t w\sigma(f) x D^2_{xf} x\right]$$
$$= 0, \tag{8.26}$$

where $D_f u$ denotes the gradient vector of u w.r.t. f and $\Delta_f u$ is the Laplacian of of u w.r.t. f and D^2_{xf} is the second derivative vector w.r.t. (x, f).

The terminal value is:

$$\mathcal{J}(T, x, f) = \frac{x^\alpha}{\alpha}.$$

Since the utility function is homogeneous, and the dynamics of the wealth process depends linearly on the control w, a logarithmic transformation can be introduced and the value function can be searched from the form:

$$\mathcal{J}(t, x, f) = \frac{x^\alpha}{\alpha}\exp\left[-\varphi(t, f)\right].$$

Therefore, we have:

$$\frac{\partial \mathcal{J}}{\partial t} = -\frac{\partial \varphi}{\partial t}\mathcal{J};\ x\frac{\partial \mathcal{J}}{\partial x} = \alpha\mathcal{J};\ x^2\frac{\partial^2 \mathcal{J}}{\partial t^2} = \alpha(\alpha - 1)\mathcal{J},$$

$$D_f \mathcal{J} = -\mathcal{J}\ D\varphi;\ \Delta_f \mathcal{J} = \left[-\Delta\varphi + D^t\varphi D\varphi\right]\mathcal{J};\ xD^2_{xf}x = -\alpha\mathcal{J}D\mathcal{J} \tag{8.27}$$

Thus, the function φ is the solution of the following semilinear PDE:

$$-\frac{\partial \varphi}{\partial t} - \frac{1}{2}\Delta\varphi + H(f, D\varphi) = 0 \text{ with } \varphi(T, f) = 0, \tag{8.28}$$

and the function H is defined by:

$$H(f,p) =$$
$$\frac{1}{2}\,||f||^2 - \ {}^t p\eta(f) + \alpha r + \max\left[\alpha^t w\,(\mu(f) - \sigma(f)p)\,x - \frac{\alpha(1-\alpha)}{2}\,\left|\left|\Sigma(f)^t w\right|\right|^2\right].$$

PROPOSITION 8.1 Equivalence of both PDE

Assume that there exists a solution φ to the previous semilinear PDE (which is supposed to be continuously differentiable w.r.t. current time t, and twice-continuously differentiable w.r.t.(x,f), with terminal value $\varphi(T,f)=0$.

Then, the value function of the first problem is given by:

$$\mathcal{J}(t,x,f) = \frac{x^\alpha}{\alpha}\exp\left[-\varphi(t,f)\right].$$

In addition, an optimal portfolio is given by the Markov control

$$\widetilde{w}_t = \widetilde{w}(t, F_t),$$

with

$$\widetilde{w}(t,f) \in \arg\min_{w\in\mathcal{A}}\left[\frac{1-\alpha}{2}\left(\left|\left|\Sigma(f)\ {}^t w\right|\right|^2\right) - \ {}^t w\,(\mu(f) - \sigma(y)D\varphi(t,f))\right].$$

If there is no specific constraint on w, the optimal Markov control is given by:

$$\widetilde{w}(t,f) = \frac{1-\alpha}{2}\left(\left|\left|\Sigma(f)^t w\right|\right|^2\right) - \ {}^t w\,(\mu(f) - \sigma(y)D\varphi(t,f)).$$

REMARK 8.2 In the case of constant coefficients, which corresponds to an extension of Merton's model, the solution does not depend on y and is equal to:

$$\varphi(t,y) = \lambda(T-t),$$

where the parameter λ is given by:

$$\lambda = \alpha r + \max_{w\in\mathcal{A}}\left[\alpha w^t \mu(y) - \frac{\alpha\,(1-\alpha)}{2}\left(\left|\left|{}^t\Sigma w\right|\right|^2\right)\right].$$

The optimal proportion is constant given by:

$$\widetilde{w} \in \arg\min_{w\in\mathcal{A}}\left[\frac{(1-\alpha)}{2}\left|\left|\Sigma(y)^t w\right|\right|^2 - \ w^t\mu(y)\right].$$

However, in a general stochastic volatility model, no closed-form solution is available for the parabolic Cauchy problem (8.28).
It must be solved numerically but is simpler that the initial Bellman equation.
□

Example 8.4 **Hull-White and Scott models**

Assume that

$$\frac{dS_t}{S_t} = (r + \widetilde{\mu}(F_t))\,dt + \rho e^{\gamma F_t} dW_{1,t} + \sqrt{1-\rho^2} e^{\gamma F_t} dW_{2,t},$$

$$dF_t = (a - \theta F_t)\,dt + dW_{2,t},$$

where a and $\gamma \neq 0$ are constant and θ is the rate of mean-reversion.

The parameter ρ is a constant correlation coefficient.

In that case, we have:

$$\eta(f) = a - \theta f, \sigma(f) = \rho e^{\gamma f} \text{ and } \Sigma^t\Sigma(f) = e^{2\gamma f}.$$

Therefore, without constraint on weighting w, the value function of the optimization problem is given by

$$\mathcal{J}(t,x,f) = \frac{x^\alpha}{\alpha} \exp\left[-\varphi(t,f)\right],$$

where $\varphi(.,.)$ is the unique solution of the previous Cauchy problem:

$$\begin{aligned} &-\frac{\partial \varphi}{\partial t} - \frac{1}{2}\frac{\partial^2 \varphi}{\partial f^2} + \frac{1}{2}\left(1 + \frac{\alpha}{1-\alpha}\rho^2\right)\left(\frac{\partial \varphi}{\partial f}\right)^2 \\ &-\left(a - \theta f + \frac{\alpha}{1-\alpha}\frac{\rho\widetilde{\mu}(f)}{e^{\gamma f}}\right)\frac{\partial \varphi}{\partial f} + \alpha r + \frac{\alpha}{1-\alpha}\frac{\mu^2(f)}{e^{2\gamma f}} \\ = {} & 0. \end{aligned} \tag{8.29}$$

Additionally, the optimal portfolio is given by:

$$\widetilde{w}_t = \frac{e^{-\gamma F_t}}{1-\alpha}\left[\frac{\widetilde{\mu}\left(F_t\right)}{e^{\gamma F_t}} - \rho\frac{\partial \varphi}{\partial f}(t, F_t)\right], \text{ a.s., } 0 \le t \le T.$$

$\square$

8.2 Optimization with constraints

As seen in Chapter 3, specific constraints are often introduced, such as no short-selling, lower and upper bounds on portfolio weights, *etc.*

In the continuous-time setting, Cvitanic and Karatzas [137] study the stochastic dynamic control problem associated to the maximization of the expected utility of consumption and terminal value.

Mathematically speaking, the set of constraints is supposed to be a given closed and convex subset of $\mathbb{R}^d$. As seen in what follows, the idea is to embed the constrained problem in a family of unconstrained ones. Then, to find an element of this family which satisfies the required constraints. Such results cover, in particular, incompleteness and no short-selling constraints. Again, this approach is based on martingale theory, duality theory, and convex analysis.

8.2.1 General result

- *The financial market* contains a riskless asset with rate r. The price S of the d risky assets is assumed to be a semimartingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ solution of the SDE:

$$dS_t = diag(S_t). \left[\mu\left(t\right) dt + \sigma\left(t\right) dW_t\right], \tag{8.30}$$

where W is a $d-$dimensional standard Brownian motion, and usual previous assumptions on coefficient functions are made.

This financial market is complete. Thus, there exists one and only one risk-neutral probability $\mathbb{Q}$ defined as follows:

Denote the relative risk process η by:

$$\eta(t) = \sigma\left(t\right)^{-1}\left[\mu\left(t\right) - r(t)\mathbb{I}\right],$$

with

$$\mathbb{E}\left[\int_0^T \left|\left|\eta(t)\right|\right|^2 dt\right] < \infty.$$

The exponential local martingale L:

$$L(t) = \exp\left[-\frac{1}{2}\int_0^t \left|\left|\eta(s)\right|\right|^2 ds - \int_0^t \eta(s) dW_s\right],$$

is the Radon-Nikodym density of the risk-neutral probability $\mathbb{Q}$ w.r.t. the probability $\mathbb{P}$.

Denote R as the discount factor:

$$R(t) = \exp\left[-\int_0^t r(s)ds\right].$$

Denote M the product: $M(t) = R(t)L(t)$.

- *Set K of constraints:* The constraints are assumed to be such that the set K is a non-empty, closed, and convex subset of $\mathbb{R}^d$.

Denote by δ the support function of the convex set $-K$, defined on $\mathbb{R}^d$, and with values in $\mathbb{R}^d \cup \{+\infty\}$:

$$\delta(x) = \delta(x, K) = \sup_{w \in K} (- \,{}^t w.x).$$

The function δ is a closed, positively homogeneous, proper convex function on $\mathbb{R}^d$ (see for example, Rockafellar [426] for details about these notions).

The effective domain of the function δ is the set $\widetilde{K}$ defined by:

$$\begin{aligned}\widetilde{K} &= \left\{x \in \mathbb{R}^d, \delta(x) < \infty\right\}, \\ &= \left\{x \in \mathbb{R}^d, \exists \beta \in \mathbb{R}, - \,{}^t w.x \leq \beta, \forall w \in K\right\}.\end{aligned}$$

The set $\widetilde{K}$ is a convex cone, called the "barrier cone" of $-K$.

In what follows, it is assumed that the function $\delta(., K)$ is continuous on $\widetilde{K}$ and bounded below on $\mathbb{R}^d$. For some $\delta_0 \in \mathbb{R}$, $\delta(x, K) \geq \delta_0$ (for example, if K contains the origin, $\delta_0 = 0$ is convenient).

Note also that the function $\delta(.)$ is subadditive:

$$\delta(x + y) \leq \delta(x) + \delta(y).$$

- *Utility functions:* these functions satisfy all the same assumptions as shown in the previous section devoted to incomplete markets. In particular, the conjugate function of a utility function U is still denoted by $\widehat{U}$:

$$\widehat{U}(y) = \max_{x>0} \left[U(y) - xy\right],\ y > 0, \tag{8.31}$$

and the inverse of the derivative marginal utility U' is denoted by J.

- *Constrained optimization problem:* The investor searchs to maximize

$$\mathbb{E}\left[\int_0^T U(t, c_t)dt + \widetilde{U}(V_T)\right],$$

on the set $\mathcal{A}_K(v_0)$ of all admissible strategies (c, w) satisfying usual assumptions (see Chapter 6):

$$\mathcal{A}_K(v_0) = \{(c, w),\ w(\omega, t) \in K,\ \text{for } \mathbb{P} \otimes dt \text{ a.s. } (\omega, t)\}.$$

The value function $\mathcal{J}_K$ is defined by:

$$\mathcal{J}_K = \sup_{(c,w)\in\mathcal{A}_K(v_0)} \mathbb{E}\left[\int_0^T U(t, c_t)dt + \widetilde{U}(V_T)\right].$$

- *When $K = \mathbb{R}^d$ (no constraint)*, recall that the solution is given by relation (6.83):

$$c_t^* = J(\lambda^*(\kappa(t))) \text{ and } V_T^* = \widetilde{J}(\lambda^*(\kappa(T))),$$

where $\kappa(t) = R(t)L(t)$.

The Lagrange parameter λ^* is determined from the budget equation. Its existence is deduced from assumptions (U) and (V), since the function F, defined by

$$F(y) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T \kappa(t)J(y\kappa(t), t)dt + \kappa(T)\widetilde{J}\left(y\kappa(T)\right)\right],$$

is continuous and non-increasing on $[0, a]$ where $a = \inf\{y\,|F(y) = 0\}$. It satisfies also:

$$\lim_{y\to 0} F(y) = +\infty; \quad \lim_{y\to+\infty} F(y) = 0.$$

Then, the function F has an inverse F^{-1}.

Define the function G by:

$$G(y) = H(F^{-1}(y)),$$

where the function H is given by:

$$H(y) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T U(J(y\kappa(t)), t)dt + \widetilde{U}\left[\widetilde{J}\left(y\kappa(T)\right)\right]\right].$$

Under assumptions (U) and (V) for utility functions U and $\widetilde{U}$, there exists an optimal strategy $(c^*, w^*) \in \mathcal{A}(v_0)$ such that:

$$\mathcal{J}(c^*, w^*, v_0) = \max_{c,w\in\mathcal{A}(v_0)} \mathcal{J}(c, w, v_0) \tag{8.32}$$

$$\text{where } \mathcal{J}(c, w, v_0) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T U(c_s, s)ds + \widetilde{U}\left(V_T^{c,w}\right)\right]. \tag{8.33}$$

Using the previous function F and the inverse functions J and $\widetilde{J}$ of marginal utility functions U' and $\widetilde{U}'$, we have:

$$\begin{aligned} c_t^* &= J(F^{-1}(v_0)\kappa(t)), \\ V_T^* &= \widetilde{J}(F^{-1}(v_0)\kappa(T)), \end{aligned}$$

and the optimal weighting w^* is deduced from the martingale representation 6.78.

- *Auxiliary unconstrained optimization problem:* Cvitanic and Karatzas [137] introduce a family of unconstrained optimization problems which embeds the constrained problem. For this purpose:

- Consider the space $\mathcal{H}$ of $\mathcal{F}_t$-progressively measurable processes $(v_t)_t$ with values in $\mathbb{R}^d$, and such that:

$$||v||^2 = \mathbb{E}\left[\int_0^T ||v_t||^2\, dt\right] < \infty.$$

- Introduce the class $\mathcal{D}$ of processes such that:

$$\mathcal{D} = \left\{\nu \in \mathcal{H},\ \int_0^T \delta\left(v_t\right) dt \leq \infty\right\}, \tag{8.34}$$

  where $\delta\left(.\right)$ is the support function of the set of constraints K. Note that:

$$\nu \in \mathcal{D} \Longleftrightarrow \nu(\omega, t) \in \widetilde{K},\ \text{ for } \mathbb{P} \otimes dt \text{ a.s. } (\omega, t), \tag{8.35}$$

  where $\widetilde{K}$ is the barrier cone of K.

For any given $\nu \in \mathcal{D}$, consider a new financial market $\mathcal{M}$, with one bond and d stocks:

$$dB_t^{(v)} = B_t^{(v)}.\left[r(t) + \delta\left(v_t\right)\right] dt, \tag{8.36}$$

$$dS_{i,t}^{(v)} = S_{i,t}^{(v)}.\left[\mu_i(t) + v_i(t) + \delta\left(v_t\right) + \sum_{j=1}^{d} \sigma_{i,j}(t) dW_{j,t}\right]. \tag{8.37}$$

Associated to the process $\nu \in \mathcal{D}$, denote also:

* The relative risk process $\eta^{(v)}$ by:

$$\eta^{(v)}(t) = \sigma\left(t\right)^{-1}\left[\mu\left(t\right) + v(t) + \delta\left(v_t\right)\mathbb{I} - \left(r(t) + \delta\left(v_t\right)\right)\mathbb{I}\right] = \eta(t) + \sigma^{-1}\nu(t). \tag{8.38}$$

* The exponential local martingale $L^{(v)}$:

$$L^{(v)}(t) = \exp\left[-\frac{1}{2}\int_0^t \left|\left|\eta^{(v)}(s)\right|\right|^2 ds - \int_0^t \eta^{(v)}(s)dW_s\right], \tag{8.39}$$

is the Radon-Nikodym density of the risk-neutral probability $\mathbb{Q}^{(v)}$ w.r.t. the probability $\mathbb{P}$.

* Denote $R^{(v)}$ the discount factor:

$$R^{(v)}(t) = \exp\left[-\int_0^t (r(s) + \delta(v_s))\, ds\right]. \tag{8.40}$$

* Denote $M^{(v)}$ the product: $M^{(v)}(t) = R^{(v)}(t)L^{(v)}(t)$.

* Consider the new set of admissible strategies:

- The wealth process $V^{(v)}$ satisfies:

$$dV_t^{(v)} = .\left[[r(t) + \delta(v_t)]\, V_t^{(v)} - c(t)\, dt\right] + .V_t^{(v)}\left[{}^t w(t)\sigma(t)dW_t^{(v)}\right], \tag{8.41}$$

where $W_t^{(v)} = W_t + \int_0^t \eta^{(v)}(s)ds$ is a standard Brownian motion under the risk-neutral probability $\mathbb{Q}^{(v)}$.

- The investor searchs to maximize (through the strategy $(c^{(v)}, w^{(v)})$):

$$\mathbb{E}\left[\int_0^T U(t, c_t^{(v)})dt + \widetilde{U}(V_T^{(v)})\right], \tag{8.42}$$

on the set $\mathcal{A}^{(v)}$ of all admissible strategies $(c^{(v)}, w^{(v)})$:

$$\mathcal{A}^{(v)}(v_0) = \left\{(c^{(v)}, w^{(v)}),\ w^{(v)}(\omega, t) \in K,\ \text{for } \mathbb{P}\otimes dt \text{ a.s. } (\omega, t)\right\}. \tag{8.43}$$

The value function $\mathcal{J}$ is still defined by:

$$\mathcal{J}^{(v)} = \sup_{(c^{(v)}, w^{(v)})\in\mathcal{A}^{(v)}} \mathbb{E}\left[\int_0^T U(t, c_t^{(v)})dt + \widetilde{U}(V_T^{(v)})\right]. \tag{8.44}$$

The Lagrange parameter $\lambda^{(v)*}$ is determined from the budget equation. Its existence is from assumptions (U) and (V). Indeed, the function $F^{(v)}$, defined by

$$F^{(v)}(y) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T \kappa^{(v)}(t)J(y\kappa^{(v)}(t), t)dt + \kappa^{(v)}(T)\widetilde{J}\left(y\kappa^{(v)}(T)\right)\right], \tag{8.45}$$

has an inverse $F^{(\upsilon)-1}$.

Define the function $G^{(\upsilon)}$ by:

$$G^{(\upsilon)}(y) = H^{(\upsilon)}(F^{(\upsilon)-1}(y)),$$

where the function $H^{(\upsilon)}$ is given by:

$$H^{(\upsilon)}(y) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T U(J(y\kappa^{(\upsilon)}(t)), t)dt + \widetilde{U}\left[\widetilde{J}\left(y\kappa^{(\upsilon)}(T)\right)\right]\right]. \tag{8.46}$$

PROPOSITION 8.2
Under assumptions (U) and (V) for utility functions U and $\widetilde{U}$, there exists an optimal strategy $(c^{(\upsilon)*}, w^{(\upsilon)*}) \in \mathcal{A}^{(\upsilon)}(v_0)$ such that:

$$\mathcal{J}(c^{(\upsilon)*}, w^{(\upsilon)*}, v_0) = \max_{c^{(\upsilon)}, w^{(\upsilon)} \in \mathcal{A}^{(\upsilon)}(v_0)} \mathcal{J}(c^{(\upsilon)}, w^{(\upsilon)}, v_0) \tag{8.47}$$

$$\text{where } \mathcal{J}(c^{(\upsilon)}, w^{(\upsilon)}, v_0) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T U(c_s^{(\upsilon)}, s)ds + \widetilde{U}\left(V_T^{(\upsilon)c,w}\right)\right]. \tag{8.48}$$

Using the previous function $F^{(\upsilon)}$ and the inverse functions J and $\widetilde{J}$ of marginal utility functions U' and $\widetilde{U}'$, we have:

$$\begin{aligned} c_t^{(\upsilon)*} &= J(F^{(\upsilon)-1}(v_0)\kappa^{(\upsilon)}(t)), \\ V_T^{(\upsilon)*} &= \widetilde{J}(F^{(\upsilon)-1}(v_0)\kappa^{(\upsilon)}(T)), \end{aligned}$$

and the optimal weighting $w^{(\upsilon)*}$ is deduced from the martingale representation (6.78).

Introduce the class $\mathcal{D}'$ of processes:

$$\mathcal{D}' = \left\{\nu \in \mathcal{D},\ F^{(\upsilon)}(y) < \infty,\ \text{for all } y \text{ a.s;}\right\}, \tag{8.49}$$

$$\mathcal{A}^{(\upsilon)}(v_0) = \{(c, w),\ w(\omega, t) \in K,\ \text{for } \mathbb{P} \otimes dt \text{ a.s. } (\omega, t)\}. \tag{8.50}$$

- The optimization problem:

$$\mathcal{J}^{(\upsilon)}(c^{(\upsilon)*}, w^{(\upsilon)*}, v_0) = \max_{c^{(\upsilon)}, w^{(\upsilon)} \in \mathcal{A}^{(\upsilon)}(v_0)} \mathcal{J}^{(\upsilon)}(c^{(\upsilon)}, w^{(\upsilon)}, v_0) \tag{8.51}$$

$$\text{with } \mathcal{J}^{(\upsilon)}(c^{(\upsilon)}, w^{(\upsilon)}, v_0) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T U(c_s^{(\upsilon)}, s)ds + \widetilde{U}\left(V_T^{(\upsilon)c^{(\upsilon)},w^{(\upsilon)}}\right)\right] \tag{8.52}$$

Then, under assumptions (U) and (V) for utility functions U and $\widetilde{U}$, there exists an optimal strategy $(c^{(\upsilon)*}, w^{(\upsilon)*}) \in \mathcal{A}^{(\upsilon)}(v_0)$ such that,

Using previous function F and the inverse functions J and $\widetilde{J}$ of marginal utility functions U' and $\widetilde{U}'$, we have:

$$
\begin{aligned}
c_t^{(\upsilon)*} &= J(F^{(\upsilon)-1}(v_0)\kappa^{(\upsilon)}(t)), \\
V_T^{(\upsilon)*} &= \widetilde{J}(F^{(\upsilon)-1}(v_0)\kappa^{(\upsilon)}(T)).
\end{aligned}
$$

- *Equivalent optimization conditions:*

PROPOSITION 8.3

Suppose that for some process $\lambda \in \mathcal{D}'$, we have:

$$
w_t^{(\lambda)} \in K,\ \delta\left(\lambda(\omega, t)\right) + {}^t w_t^{(\lambda)} \lambda(\omega, t) = 0. \tag{8.53}
$$

Then the pair $\left(c^{(\lambda)*}, w^{(\lambda)*}\right)$ belongs to $\mathcal{A}^{(\lambda)}(v_0)$ and is optimal for the constrained optimization problem in the original market.

- Consider a solution (c_K^*, w_K^*) of the constrained optimization problem (A):

$$
\sup_{(c_K, w_K) \in \mathcal{A}_K(v_0)} \mathbb{E}\left[\int_0^T U(t, c_{K,t})dt + \widetilde{U}(V_{K,T})\right].
$$

Cvitanic and Karatzas [137] characterize the solution of Problem (A) by using the following conditions (B)-(E) for a given process λ in the class $\mathcal{D}'$:

* *Financibility of* $\left(c^{(\lambda)}, V_T^{(\lambda)}\right)$:

There exists a portfolio process w_K such that $(c^{(\lambda)}, w^{(\lambda)}) \in \mathcal{A}_K$ and:

$$
\begin{aligned}
w_t^{(\lambda)}(\omega, t) \in K,\ \delta\left(\lambda(\omega, t)\right) + {}^t w_t^{(\lambda)} \lambda(\omega, t) = 0. \\
V^{(v_0, c^{(\lambda)}, w^{(\lambda)})}(\omega, t) = V^{(\lambda)}(\omega, t) \text{for } \mathbb{P} \otimes dt \text{ a.s. } (\omega, t).
\end{aligned}
$$

* *Minimality of* λ. For every $\nu \in \mathcal{D}$, we have:

$$
\mathbb{E}\left[\int_0^T U(t, c_t^{(\lambda)})dt + \widetilde{U}(V_T^{(\lambda)})\right] \leq \mathbb{E}\left[\int_0^T U(t, c_t^{(\nu)})dt + \widetilde{U}(V_T^{(\nu)})\right].
$$

* *Dual optimality of* λ. For every $\nu \in \mathcal{D}$, we have:

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\left[\int_0^T \kappa^{(\upsilon)}(t) J(y\kappa^{(\upsilon)}(t), t)dt + \kappa^{(\upsilon)}(T)\widetilde{J}\left(y\kappa^{(\upsilon)}(T)\right)\right] \\
\leq &\mathbb{E}_{\mathbb{P}}\left[\int_0^T \kappa^{(\upsilon)}(t) J(y\kappa^{(\lambda)}(t), t)dt + \kappa^{(\upsilon)}(T)\widetilde{J}\left(y\kappa^{(\lambda)}(T)\right)\right].
\end{aligned}
$$

* *Parsimony of* λ. For every $\nu \in \mathcal{D}$, we have:

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^T \kappa^{(\upsilon)}(t))c_t^{(\lambda)})dt + \kappa^{(\upsilon)}(T)V_T^{(\lambda)}\right] \leq v_0.$$

THEOREM 8.4 **Equivalence of conditions**
Conditions (B)-(E) are equivalent and imply property (A) with

$$(c_K^*, w_K^*) = (c^{(\lambda)}, w^{(\lambda)}). \tag{8.54}$$

In addition, conversely, property (A) implies the existence of $\lambda \in \mathcal{D}'$ *which satisfies conditions (B)-(E) with* $w_K^* = w^{(\lambda)}$ *under the following assumptions:*

(i) The utility functions satisfy:

$$\begin{cases} c \to cU'(c) \\ v \to v\widetilde{U}'(v) \end{cases} \text{are nondecreasing on } (0,\infty), \tag{8.55}$$

and for some $\alpha \in (0,1), \gamma \in (1,\infty)$, *we have:*

$$\alpha U'(x) \geq U'(\gamma x), \text{ for all } x > 0. \tag{8.56}$$

From the previous result, we are led again to the dual stochastic control problem:

$$\mathcal{J}(y) = \inf_{\nu \in \mathcal{D}} \mathbb{E}\left[\int_0^T \widehat{U}\left(t, y\kappa^{(\nu)}(t)\right)dt + \widehat{\widetilde{U}}\left(y\kappa^{(\nu)}(T)\right)\right]. \tag{8.57}$$

THEOREM 8.5 **Existence of a solution of the constrained optimization**
Under all previous assumptions on the utility functions, there exists an optimal pair (c_K^*, w_K^*) *for the constrained portfolio.*

8.2.2 Basic examples

Examine the optimal solution for various constraints: no short-selling, upper and lower bounds on the weighting vector, *etc.*

8.2.2.1 Standard strategy constraints

Example 8.5 **General closed convex cone K**
We have:

$$\delta(x) = \begin{cases} 0 \text{ if } x \in K \\ \infty \text{ if } x \in \widetilde{K} \end{cases}.$$

Recall that $\widetilde{K}$ is the polar cone of $-K$:

$$\widetilde{K} = \left\{x \in \mathbb{R}^d, {}^t wx \geq 0, \forall w \in K\right\}.$$

Note that, for the unconstrained case, $K = \mathbb{R}^d$,

$$\delta(x) = \begin{cases} 0 \text{ if } x = 0 \\ \infty \text{ otherwise} \end{cases}, \text{ with } \widetilde{K} = \{0\}.$$

□

Example 8.6 **No short-selling**
Let:

$$K = \left\{w \in \mathbb{R}^d, w_i \geq 0, \forall i = 1, ..., d\right\}.$$

Then:

$$\delta(x) = \begin{cases} 0 \text{ if } x \in K \\ \infty \text{ if } x \notin K \end{cases} \text{ and } \widetilde{K} = K.$$

□

Example 8.7 **Incomplete market**
Let:

$$K = \left\{w \in \mathbb{R}^d, w_i = 0, \forall i = m+1, ..., d\right\}, \text{ for some } m \in \{1, ..., d-1\}.$$

Then:

$$\delta(x) = \begin{cases} 0, \ x_1 = ... = x_m = 0, \\ \infty, \text{ otherwise.} \end{cases}$$

and

$$\widetilde{K} = \left\{x \in \mathbb{R}^d, x_i = 0, \forall i = 1, ..., m\right\}.$$

This case corresponds to an incomplete financial market driven by a multi-dimensional Brownian motion, which contains m stocks with $m < d$.

If, furthermore, there are no short-selling conditions then:

$$K = \left\{w \in \mathbb{R}^d, w_i \geq 0, \forall i \leq d \text{ and } w_i = 0, \forall i = m+1, ..., d\right\},$$

$$\delta(x) = \begin{cases} 0, \ (x_1, ..., x_m) \in [0, \infty[^m \\ \infty, \text{ otherwise,} \end{cases}$$

and

$$\widetilde{K} = \left\{x \in \mathbb{R}^d, x_i \geq 0, \forall i = 1, ..., m\right\}.$$

□

Example 8.8 **Rectangular constraints**
Let:

$$K = \prod_{i=1}^{d} K_i \text{ with } K_i = [\alpha_i, \beta_i],$$

for some fixed numbers: $-\infty \leq \alpha_i \leq \beta_i \leq +\infty$.

Then:

$$\delta(x) = \sum_{i=1}^{d} \beta_i x_i^- - \sum_{i=1}^{d} \alpha_i x_i^+,$$

and

$$\widetilde{K} = \left\{ x \in \mathbb{R}^d, x_i \geq 0, \forall i \in \mathcal{S}^+ \text{ and } x_i \leq 0, \forall i \in \mathcal{S}^- \right\},$$

where

$$\begin{aligned} \mathcal{S}^+ &= \{ i = 1, ..., d \,|\, \beta_i = +\infty \}, \\ \mathcal{S}^- &= \{ j = 1, ..., d \,|\, \alpha_j = -\infty \}. \end{aligned}$$

$\square$

8.2.2.2 Logarithmic utility case

Assume that $U(c) = \ln c$ and $\widetilde{U}(v) = \ln v$. Then, we have: for every $\nu \in \mathcal{D}$,

$$c^{(\upsilon)}(t) = \frac{v_0}{T+1} \frac{1}{\kappa^{(\upsilon)}(t)}; V^{(\upsilon)}(T) = \frac{v_0}{T+1} \frac{1}{\kappa^{(\upsilon)}(T)}. \tag{8.58}$$

Example 8.9
Note that $\mathcal{D} = \mathcal{D}'$. Thus:

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}} \left[\int_0^T \kappa^{(\upsilon)}(t) J(y\kappa^{(\lambda)}(t), t) dt + \kappa^{(\upsilon)}(T) \widetilde{J} \left(y\kappa^{(\lambda)}(T) \right) \right] \\ &= -(1+T) \left(1 + \ln \frac{1+T}{v_0} \right) + \mathbb{E}_{\mathbb{P}} \left[\ln \frac{1}{\kappa^{(\upsilon)}(T)} + \int_0^T \ln \frac{1}{\kappa^{(\upsilon)}(t)} dt \right], \end{aligned}$$

and,

$$\mathbb{E}_{\mathbb{P}} \left[\ln \frac{1}{\kappa^{(\upsilon)}(t)} \right] = \mathbb{E}_{\mathbb{P}} \left[\int_0^t \left(r(s) + \delta(\nu(s)) + \frac{1}{2} \left|\left| \eta(s) + \sigma^{-1}(s) v(s) \right|\right|^2 \right) \right] ds.$$

Therefore, the optimization problem is equivalent to a pointwise minimization of the convex function $\delta(x) + \frac{1}{2} \left|\left| \eta(t) + \sigma^{-1}(t) x \right|\right|^2$ over $x \in \widetilde{K}$, for every $t \in [0, T]$.

Thus the process λ is determined by:

$$\lambda(t) = \arg \min_{x \in \widetilde{K}} \left[2\delta(x) + \left|\left| \eta(t) + \sigma^{-1}(t) x \right|\right|^2 \right]. \tag{8.59}$$

Finally, we deduce:

$$w_K(t) = \sigma(t)^t \sigma(t)^{-1} \left[\lambda(t) + \mu(t) - r(t)\mathbb{I}\right],$$
$$c_K(t) = \frac{v_0}{T+1} \frac{1}{\kappa^{(v)}(t)} = \frac{V^{(\lambda)}(t)}{1+(T-t)}.$$

∎

For all previous examples corresponding to various constraint sets K, we have $\delta(.) = 0$ on $\widetilde{K}$. Thus, for the logarithmic case, the problem of determining the process $\lambda \in \mathcal{D}$ reduces to that of minimizing pointwise a simple quadratic form, over $\widetilde{K}$:

$$\lambda(t) = \arg\min_{x \in \widetilde{K}} \left[2\delta(x) + \left|\left|\eta(t) + \sigma^{-1}(t)x\right|\right|^2\right]. \tag{8.60}$$

For the unconstrained case, $\lambda(t) = 0$ and we recover the standard solution:

$$w_K(t) = \sigma(t)^t \sigma(t)^{-1} \left[\lambda(t) + \mu(t) - r(t)\mathbb{I}\right].$$

Therefore, for the logarithmic case, we get explicit formulas:

- For the incomplete case, set:

$$\sigma(t) = \begin{bmatrix} \Sigma(t) \\ \rho(t) \end{bmatrix},$$

  where $\Sigma(\omega, t)$ is an $(m \times d)$ matrix of full rank and $\rho(\omega, t)$ is an $(n \times d)$ matrix with orthogonal rows that span the kernel of $\Sigma(\omega, t)$ for every (ω, t) (we have: $\rho(\omega, t)^t \rho(\omega, t) = I_n$ and $\Sigma(\omega, t)^t \rho(\omega, t) = 0$ and $n = d - m$).

  Then, set:

$$M(t) = {}^t\left(\mu_1(t), ..., \mu_m(t)\right),$$
$$a(t) = {}^t\left(b_{m+1}(t), ..., b_d(t)\right),$$
$$\Lambda(t) = {}^t\Sigma(t)\Sigma(t)^t\Sigma(t)^{-1}\left[B(t) - r(t)\mathbb{I}\right].$$

  We have:

$$\eta(t) = \Lambda(t) + {}^t\rho(t)\left[M(t) - r(t)\mathbb{I}\right].$$

  For any $v \in \widetilde{K}$, necessarily of the form $\nu(t) = \begin{bmatrix} 0_m \\ N \end{bmatrix}$ for some $N \in \mathbb{R}^m$,

$$\left|\left|\eta(t) + \sigma^{-1}(t)\nu\right|\right|^2 = \left|\left|\Lambda(t) + {}^t\rho(t)\left(a(t) - r(t)\mathbb{I}_n + \mathbb{N}\right)\right|\right|^2.$$

Since $\Lambda(t)$ and ${}^t\rho(t)\,(a(t) - r(t)\mathbb{I}_n + \mathbb{N})$ are orthogonal, we have:

$$\left|\left|\eta(t) + \sigma^{-1}(t)\nu\right|\right|^2 = \left|\left|\Lambda(t)\right|\right|^2 + \left|\left|{}^t\rho(t)\,(a(t) - r(t)\mathbb{I}_n + \mathbb{N})\right|\right|^2.$$

Therefore, the minimization 8.60 is achieved by the random vector,

$$\lambda(t) = \begin{bmatrix} 0_m \\ \Lambda(t) \end{bmatrix} \text{ where } \Lambda(t) = r(t)\mathbb{I}_n - a(t).$$

Thus, we deduce (result in Karatzas et al. [319]):

$$w_K(t) = \begin{bmatrix} \Sigma(t)^t\Sigma(t)^{-1}\left[M(t) - r(t)\mathbb{I}_m\right] \\ 0_n \end{bmatrix}.$$

- For the rectangular constraints, the process λ can be also determined:

  - First case (only one stock): $\mathcal{S}^+ = \mathcal{S}^- = \varnothing$.

  Since $d = 1$, K has the form $[\alpha, \beta]$ and $\delta(x) = \beta x^- - \alpha x^+$. The process λ is defined by:

$$\lambda(t) = \begin{cases} \sigma(t)\left[\sigma(t)\beta - \eta(t)\right], \text{ if } \sigma(t)\beta < \eta(t), \\ \sigma(t)\left[\sigma(t)\alpha - \eta(t)\right], \text{ if } \sigma(t)\alpha > \eta(t), \\ 0, \text{ otherwise.} \end{cases}$$

  Therefore, the optimal portfolio is given by:

$$w_K(t) = \begin{bmatrix} \beta \text{ if } \sigma^{-1}(t)\eta(t) > \beta, \\ \alpha \text{ if } \sigma^{-1}(t)\eta(t) < \alpha, \\ \sigma^{-1}(t)\eta(t), \text{ otherwise.} \end{bmatrix}$$

  Note that the optimal portfolio $w_K(t)$ is equal to the unconstrained optimal one $w_{\mathbb{R}^d}(t)$, as long as $w_{\mathbb{R}^d}(t)$ is in the interval $[\alpha, \beta]$. Otherwise, $w_K(t)$ is equal to the closest point to $w_{\mathbb{R}^d}(t)$.

  - Second case (two stocks): suppose for example that $\alpha = {}^t(0, 0)$, $\beta = {}^t(1, 1)$, $\eta = {}^t(1, 2)$ and

$$\sigma(t) = \begin{bmatrix} 1 & -10 \\ -1 & 1 \end{bmatrix}.$$

  For the unconstrained case, the optimal portfolio is given by:

$$w_{\mathbb{R}^d}(t) = \sigma(t)^{-1}\eta = {}^t(-1/3, -4/3).$$

  The optimal constrained portfolio $w_K(t)$ is no longer the portfolio in K which is the closest one to the unconstrained optimal portfolio $w_{\mathbb{R}^d}(t)$. Otherwise, it would be equal to ${}^t(0, 0)$.

Indeed, the minimization:

$$\lambda(t) = \arg\min_{x\in\widetilde{K}} \left[\frac{1}{2}\left|\left|\eta+\sigma^{-1}x\right|\right|^2 - {}^t\alpha x^+ + {}^t\beta x^-\right], \; x\in\mathbb{R}^2, \qquad (8.61)$$

leads to the value $\lambda = {}^t(13.5, 0)$, and the optimal portfolio is given by:

$$w_K(t) = {}^t\sigma^{-1}\left(\eta+\sigma^{-1}\lambda\right) = {}^t(0, 1/2),$$

which means that we must not invest on the first stock and invest half on the second one.

8.2.2.3 Deterministic coefficients case

- We get feedback formulas: There exists a formal Hamilton-Jacobi-Bellman equation associated with the dual optimization problem (8.11):

$$\frac{\partial Q}{\partial t}+\inf_{x\in\widetilde{K}}\left[\frac{1}{2}\frac{\partial^2 Q}{\partial v\partial v}\left|\left|\eta(t)+\sigma^{-1}(t)x\right|\right|^2-\nu\frac{\partial Q}{\partial v}\delta(x)\right]-\nu\frac{\partial Q}{\partial v}r(t)+\widehat{U}(t,v)=0, \qquad (8.62)$$

and

$$Q(T,v)=\widehat{\widetilde{U}}(v).$$

If there exists a solution Q which satisfies mild growth conditions, then the dual function satisfies (see Fleming and Rischel [231]).

Suppose that $\delta(.) = 0$ on $\widetilde{K}$, such as for all previous constraint sets K. Thus the process

$$\lambda(t) = \arg\min_{x\in\widetilde{K}}\left[2\delta(x)+\left|\left|\eta(t)+\sigma^{-1}(t)x\right|\right|^2\right], \qquad (8.63)$$

is deterministic and constant w.r.t. v. Then Equation (8.62) becomes:

$$\frac{\partial Q}{\partial t}+\frac{1}{2}\left|\left|\eta^{(\lambda)}(t)\right|\right|^2\nu^2\frac{\partial^2 Q}{\partial v\partial v}-r(t)y\frac{\partial Q}{\partial v})+\widehat{U}(t,v)=0. \qquad (8.64)$$

Consider for instance the case $U(x)=\widetilde{U}(x)=\frac{x^\alpha}{\alpha}$. Then $\widehat{U}(x)=\widehat{\widetilde{U}}(x)=\frac{y^{-\delta}}{\delta}$ with $\delta=\frac{\alpha}{1-\alpha}$. Therefore, the Cauchy problem (8.62) has a solution of the form:

$$Q(t,v)=\frac{1}{\rho}v^{-\rho}u(t),$$

where the function $u(.)$ is the solution of the following ODE:

$$\frac{du}{dt}+h(t)u(t)+1=0, \; u(T)=1,$$

with:

$$h(t) = \rho \inf_{x \in \widetilde{K}} \left[\frac{1+\rho}{2} \left|\left|\eta(t) + \sigma^{-1}(t)x\right|\right|^2 + \delta(x) \right] + r(t)\rho.$$

The process λ is again deterministic and:

$$\lambda(t) = \arg\min_{x \in \widetilde{K}} \left[2(1-\alpha)\delta(x) + \left|\left|\eta(t) + \sigma^{-1}(t)x\right|\right|^2 \right]. \tag{8.65}$$

Cvitanic and Karatzas [137] provide a general result for the deterministic case.

PROPOSITION 8.4
Suppose that $r(.), b(.)$, and $\sigma(.)$ are deterministic and that there exists a deterministic process $\lambda(.) \in \mathcal{D}$, which achieves the infimum of dual problem (8.11), and that: for some real numbers $\alpha > 0, \beta > 0$, and $M > 0$,

$$J(t,y) + \widetilde{J}(y) + |J'(t,y)| + \left|\widetilde{J}'(y)\right| \leq M(y^{\alpha} + y^{-\beta}), 0 < y < \infty.$$

With all assumptions introduced in this section, the optimal process of consumption/investment (c^*_K, w^*_K) is given in feedback form w.r.t. the wealth $V^{(\lambda)}(t)$ by:

$$c^{(\lambda)}(t) = J(t, F^{-1}(t, V^{(\lambda)}(t))),$$
$$w_K(t) = -\sigma(t)^t \sigma(t)^{-1} \left[\lambda(t) + \mu(t) - r(t)\mathbb{I}\right] \frac{F^{-1}(t, V^{(\lambda)}(t))}{V^{(\lambda)}(t) F^{-1}(t, V^{(\lambda)}(t))}.$$

REMARK 8.3 This result shows that, for the deterministic case, the weighting ratios are still constant, despite the additional constraints on the portfolio strategies:

$$\frac{w_{K,i}(t)}{w_{K,j}(t)} = \frac{\sigma(t)^t \sigma(t)^{-1} \left[\lambda(t) + \mu(t) - r(t)\mathbb{I}\right]_i}{\sigma(t)^t \sigma(t)^{-1} \left[\lambda(t) + \mu(t) - r(t)\mathbb{I}\right]_j}.$$

$\square$

8.3 Optimization with transaction costs

As shown by Magill and Constantinides [368], Davis and Norman [149], and Shreve and Soner [470], the standard approach to utility maximization under transaction costs is based on the analytical study of the value function which usually leads to an optimal strategy corresponding to:

- No transaction in a certain region. With a single risky asset, no trade is made when the stock weight lies in a given interval.
- Minimal transactions at the boundary such that the weighting vector stays in the region.

This kind of result holds for infinite horizon: it is always optimal to invest in the stock (even a small amount) if the rate of return is positive.

For finite horizon, Cvitanic and Karatzas [139] show that the result may be quite different. If the difference between the stock return and the riskless return is non-negative but small, and the time horizon is also relatively small, then it may be optimal not to trade at all.

8.3.1 The infinite-horizon case

Consider the model introduced by Davis and Norman [149], and further studied by Shreve and Soner [470] who use the concept of viscosity solutions to Hamilton-Jacobi-Bellman equations.

- *The financial market* consists of one riskless asset, the bond B, and one risky asset, the stock S, given by:

$$dB_t = B_t r(t)dt, dS_t = S_{t_-}.\left[\mu(t)dt + \sigma(t)dW_t\right],$$

where W is a standard Brownian motion, for $t \in [0,T]$, on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$.

The processes $r(t), \mu(t)$, and $\sigma(t)$ are assumed to be measurable, $\mathcal{F}_t$-adapted, and uniformly bounded on $[0,T] \times \Omega$. Besides, the process $\sigma(t)$ is assumed to be uniformly bounded away from zero.

- *The trading strategy* is a pair (L, M) of adapted and left-continuous processes on $[0,T]$ with non-decreasing paths and $L(0) = M(0) = 0$. The process L (respectively M) denotes the cumulative purchases (resp. sales). It is the total amount transferred from bank account to stock (resp. from stock to bank account).

- *The transaction costs* are proportional : $0 < \lambda, \mu < 1$.

- *The portfolio holdings* (V_B, V_S) corresponding to a given strategy (L, U) evolve according to:

$$V_{B,t} = V_{B,0} - (1+\lambda)L_t + (1-\mu)M_t + \int_0^t V_{B,u}\,(r_u - c_u)\,du, \quad (8.66)$$

$$V_{S,t} = V_{S,0} + L_t - M_t + \int_0^t V_{S,u}\left[\mu(u)du + \sigma(u)dW_u\right]. \quad (8.67)$$

- *The solvency region* is defined by:

$$\mathcal{S}_{\lambda,\mu} = \left\{(x,y) \in \mathbb{R}^2\,|x + (1-\mu)y \geq 0 \text{ and } x + (1+\lambda)y \geq 0\right\}.$$

Denote respectively by ∂_μ^+ and ∂_λ^- the upper and lower boundaries of the solvency region. The investor's net wealth is equal to zero on $\partial_\mu^+ \cup \partial_\lambda^-$.
- *The consumption rate* is an adapted process denoted by c.
- An admissible strategy (c, L, M) is such that

$$\mathbb{P}\left[(L(t), M(t)) \in \mathcal{S}_{\lambda,\mu}, \text{ for all } t \geq 0\right] = 1. \quad (8.68)$$

- *The maximization of expected utility from consumption* is defined by:

$$\mathcal{J}(x,y) = \sup_{(c,L,M)\in\mathcal{S}_{\lambda,\mu}} \mathbb{E}_{x=V_{B,0},y=V_{S,0}}\left[\int_0^\infty e^{-\delta t}U(c_t)dt\right]. \quad (8.69)$$

8.3.1.1 A special case

Assume that the utility function is a power function: $U(x) = \frac{x^\alpha}{\alpha}$. Suppose also that both processes L and U are absolutely continuous w.r.t. the current time:

$$L_t = \int_0^t l_t dt \text{ and } M_t = \int_0^\infty m_t dt.$$

The Bellman equation to be solved for the value function $\mathcal{J}$ is

$$\max_{(c,l,m)\in\mathcal{S}_{\lambda,\mu}}$$

$$\begin{aligned}&\left[\frac{1}{2}\sigma^2 y^2 \frac{\partial^2\mathcal{J}(x,y)}{\partial y^2} + rx\frac{\partial\mathcal{J}(x,y)}{\partial x} + \mu y\frac{\partial\mathcal{J}(x,y)}{\partial y} + \frac{1}{\alpha}c^\alpha - c\frac{\partial\mathcal{J}(x,y)}{\partial x}\right]\\ &+\left[-(1+\lambda)\frac{\partial\mathcal{J}(x,y)}{\partial x} + \frac{\partial\mathcal{J}(x,y)}{\partial y}\right]l + \left[(1-\mu)\frac{\partial\mathcal{J}(x,y)}{\partial x} - \frac{\partial\mathcal{J}(x,y)}{\partial y}\right]m\\ &\qquad -\delta\mathcal{J}(x,y) = 0.\end{aligned}$$

The maximum is achieved for:

$$c = \left(\frac{\partial\mathcal{J}(x,y)}{\partial x}\right)^{1/(1-\alpha)},$$

$$l = \begin{cases} K \text{ if } \frac{\partial \mathcal{J}(x,y)}{\partial y} \geq (1+\lambda)\frac{\partial \mathcal{J}(x,y)}{\partial x} \\ 0 \text{ if } \frac{\partial \mathcal{J}(x,y)}{\partial y} < (1+\lambda)\frac{\partial \mathcal{J}(x,y)}{\partial x} \end{cases} \text{ and } m = \begin{cases} 0 \text{ if } \frac{\partial \mathcal{J}(x,y)}{\partial y} > (1-\mu)\frac{\partial \mathcal{J}(x,y)}{\partial x} \\ K \text{ if } \frac{\partial \mathcal{J}(x,y)}{\partial y} \leq (1-\mu)\frac{\partial \mathcal{J}(x,y)}{\partial x} \end{cases}.$$

Thus, the optimal strategy is bang-bang: buying and selling either are made at maximum rate, or not at all.

The solvency region splits into three parts: "buy" (B) , "sell" (S), and "no transaction" (NT).

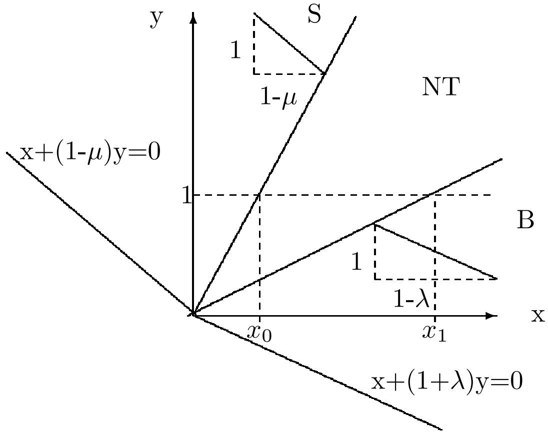


FIGURE 8.1: Bond/stock space: directions of finite transactions and solvency region

At the boundary between (B) and (NT) subsets, we have:

$$\frac{\partial \mathcal{J}(x,y)}{\partial y} = (1+\lambda)\frac{\partial \mathcal{J}(x,y)}{\partial x}.$$

At the boundary between (S) and (NT) subsets, we have:

$$\frac{\partial \mathcal{J}(x,y)}{\partial y} = (1-\mu)\frac{\partial \mathcal{J}(x,y)}{\partial x}.$$

Then, these boundaries are to be determined precisely. For this purpose, assuming that the value function satisfies an homothetic property:

$$\mathcal{J}(x,y) = y^{\alpha}\Psi\left(\frac{x}{y}\right),$$

where $\Psi(x) = \mathcal{J}(x,1)$. Therefore, $\mathcal{J}$ is constant along the lines of slope $(1+\lambda)^{-1}$ in B, and $(1-\mu)^{-1}$ in S. Thus:

$$\Psi(x) = \frac{1}{\alpha}A\left(x+1-\mu\right)^{\alpha}, \; x \le x_0,$$
$$\Psi(x) = \frac{1}{\alpha}B\left(x+1+\lambda\right)^{\alpha}, \; x \ge x_1, \tag{8.70}$$

for some constants A and B and x_0 and x_1 defined as in the previous figure.

THEOREM 8.6 *Davis and Norman [**149**]*
Assume that:
i) Well-posed condition: $\delta > \alpha\left[r + (\mu - r)^2/\sigma^2(1-\alpha)\right]$.
ii) Transaction costs: $\lambda \in [0,\infty[$, $\mu \in [0,1[$ *and max*$(\lambda,\mu) > 0$.
Then:
1) Power utility case: $(U(x) = x^{\alpha}/\alpha)$. *Suppose that:*

$$r < \mu < r + (1-\alpha)\,\sigma^2.$$

Then, the optimal solutions are defined from Equation (8.70): Let NT denote the closed wedge $\left\{(x,y) \in \mathbb{R}^{+2} \,\middle|\, \frac{1}{x_1} < \frac{y}{x} < \frac{1}{x_0}\right\}$. *For* $(x,y) \in NT - \{(0,0)\}$, *define*

$$c^*(x,y) = y\left[\Psi'\left(\frac{x}{y}\right)\right]^{1/(1-\alpha)}.$$

Then, the process $c_t^* = c^*(L_t^*, M_t^*)$, *where both processes* L_t^* *and* M_t^* *satisfy Equations 8.66 and 8.67, is optimal for any initial investment* (x,y) *in NT. If* $(x,y) \notin NT$, *then an immediate transaction leads to the closest point in NT. The processes* L_t^* *and* M_t^* *are the local times of the wealth at the boundaries of the no transaction region NT with regions B and S:*

$$L_t = \int_0^t \mathbb{I}_{\{(V_{B,s},V_{S,s})\in\partial B\}} dL_s \text{ and } M_t = \int_0^t \mathbb{I}_{\{(V_{B,s},V_{S,s})\in\partial S\}} dM_s.$$

Thus, the optimal strategy is to trade minimally in order to keep the stock weight between $(1+x_1)^{-1}$ *and* $(1+x_0)^{-1}$.

2) Logarithmic utility case: $(U(x) = \ln x)$. *Suppose that:*

$$r < \mu < r + \sigma^2.$$

There are constants $\widetilde{x}_0, \widetilde{x}_1, \widetilde{A}$ *and* $\widetilde{B}$ *such that:*

$$\widetilde{\Psi}(x) = \frac{1}{\alpha}\ln\left[\widetilde{A}\left(x+1-\mu\right)\right], \; x \le \widetilde{x}_0,$$
$$\widetilde{\Psi}(x) = \frac{1}{\alpha}\ln\left[\widetilde{B}\left(x+1+\lambda\right)\right], \; x \ge \widetilde{x}_1, \tag{8.71}$$

Define the function $\widetilde{c}^*(.,.)$ *by:*

$$\widetilde{c}^*(x,y) = y\left[\widetilde{\Psi}'\left(\frac{x}{y}\right)\right]^{-1}.$$

Note that, if $\mu = r$ *and* $\delta > \alpha r$*, then the optimal strategy is to close out any position in stock and to consume optimally from the bank account.*

8.3.2 The finite-horizon case

Consider the model introduced by Cvitanic and Karatzas [139]. The financial market consists of the same assets as in the previous case, transaction costs are still proportional and the trading strategies (L, M) are also defined as previously.

- *Introduce the following auxiliary martingales*: D denotes the class of pairs of strictly positive $(\mathcal{F}_t)$-martingales (Z_0, Z_1) with:

$$Z_0(0) = 1, Z_1(0) \in [S_0(1-\mu), S_0(1+\lambda)], \tag{8.72}$$

and

$$(1-\mu) \leq R(t) = \frac{Z_1(t)}{Z_0(t)P(t)} \leq (1+\lambda), \tag{8.73}$$

where P is the discounted price of the stock: $P(t) = \frac{S(t)}{B(t)}$.

The martingales Z_0 and $Z_1$are the feasible state-price densities for holdings in bank and stock in the markets with transaction costs. From the martingale representation theorem, there exist predictable processes θ_0 and θ_1 such that:

$$Z_0(t) = Z_0(0)\exp\left[\int_0^t \theta_0(u)dW_u - \frac{1}{2}\int_0^t \theta_0^2(u)du\right], \tag{8.74}$$

$$Z_1(t) = Z_1(0)\exp\left[\int_0^t \theta_1(u)dW_u - \frac{1}{2}\int_0^t \theta_1^2(u)du\right]. \tag{8.75}$$

Denote Z_0^c the density process when there is no transaction cost. We have:

$$Z_0^c(t) = Z_0^c(0)\exp\left[\int_0^t \theta_0^c(u)dW_u - \frac{1}{2}\int_0^t \theta_0^{c2}(u)du\right],$$

with

$$\theta_0^c(t) = \frac{r(t)-\mu(t)}{\sigma(t)}.$$

- *The utility function on wealth* $U : (0,+\infty) \to \mathbb{R}$ satisfies usual assumptions as in previous sections. The inverse of the marginal utility is again denoted by J: $J = (U')^{-1}$.

- *The maximization of expected utility from terminal wealth* is solved as follows.

The terminal wealth V_{B,T^+} is defined by:

$$V_{B,T^+} = V_{B,T} + f(V_{S,T}) \text{ with } f(u) = \begin{cases} (1+\lambda)u \text{ if } u \leq 0, \\ (1-\mu)u \text{ if } u > 0. \end{cases}$$

This means that, at the end of the management period, the investor liquidates his position on stock and transfers it to a bank account with transaction costs.

For an initial holding $V_S(0)$, we have to search for an optimal strategy (L^*, U^*) that maximizes expected utility from terminal wealth:

$$\mathcal{J}(V_{B,0}, V_{S,0}) = \sup_{(L,M)} \mathbb{E}\left[V_{B,T} + f(V_{S,T})\right]. \tag{8.76}$$

Consider the dual problem

$$\widehat{\mathcal{J}}(\zeta, V_{S,0}) = \inf_{(Z_0,Z_1)\in D} \mathbb{E}\left[\widehat{U}\left(\zeta\frac{Z_0(T)}{B(T)}\right) + \frac{V_{S,0}}{S_0}\zeta Z_1(T)\right], \tag{8.77}$$

under the assumption that there exists a pair $(Z_0^*, Z_1^*) \in D$ that achieves the infimum, for all $0 < \zeta < \infty$. Additionally, for all $0 < \zeta < \infty$, we have:

$$\widehat{\mathcal{J}}(\zeta, V_{S,0}) < \infty \text{ and } \mathbb{E}\left[\frac{Z_0^*(T)}{B(T)} J\left(\zeta\frac{Z_0^c(T)}{B(T)}\right)\right],$$

where $Z_0^c(T)$ is the density process when there is no transaction cost.

Note for example that this assumption is satisfied if $V_{S,0} = 0$, and either $U(x) = ln(x)$ or $U(x) = x^\alpha/\alpha$.

Consider the function Γ

$$\zeta \rightarrow \Gamma(\zeta) = \mathbb{E}\left[\frac{Z_0^*(T)}{B(T)} J\left(\zeta\frac{Z_0^*(T)}{B(T)}\right)\right].$$

Since the function Γ is continuous and strictly decreasing, there exists a unique ζ^* such that

$$\mathbb{E}\left[\frac{Z_0^*(T)}{B(T)} J\left(\zeta^*\frac{Z_0^*(T)}{B(T)}\right)\right] = V_{B,0} + \frac{V_{S,0}}{S_0}\mathbb{E}\left[Z_1^*(T)\right].$$

We deduce:

THEOREM 8.7

(Cvitanic and Karatzas [139]) Under previous assumptions, there exists an optimal pair (L^*, M^*) *such that:*

$$V^*_{B,T^+} = V^*_{B,T} + f(V^*_{S,T}) = J\left(\zeta^* \frac{Z^*_0(T)}{B(T)}\right), \tag{8.78}$$

with the following property:

$$\begin{aligned} &L^* \text{ is flat off the set } \{0 \le t \le T \,|\, R^*(t) = 1 + \lambda\}, \\ &M^* \text{ is flat off the set } \{0 \le t \le T \,|\, R^*(t) = 1 - \mu\}, \end{aligned}$$

and:

$$\frac{V^*_{B,t} + R^*(t)V^*_{S,t}}{B(t)} = \mathbb{E}_{\mathbb{Q}^*_0}\left[J\left(\zeta^* \frac{Z^*_0(T)}{B(T)}\right)\frac{1}{B(T)} \,|\mathcal{F}_t\right],$$

where $R^*(t) = Z^*_1(t)/Z^*_0(t)$ *and* $\mathbb{Q}^*_0$ *is defined by* $Z^*_0 = \frac{d\mathbb{Q}^*_0}{d\mathbb{P}}$.

The following examples illustrate the fact that for finite-horizon, it may be optimal not to trade at all.

Example 8.10

Assume that the rate r is deterministic and the inital amount $V_{S,0}$ invested on stock is equal to 0. In that case, we have:

$$\widehat{\mathcal{J}}(\zeta, 0) = \inf_{(Z_0,Z_1)\in D} \mathbb{E}\left[\widehat{U}\left(\zeta \frac{Z_0(T)}{B(T)}\right)\right] \ge \widehat{U}\left(\frac{\zeta}{B(T)}\right).$$

Thus, the infimum is achieved by any pair $(Z^*_0(.), Z^*_1(.))$ such that $Z^*_0(.) \equiv 1$ (then, with $(1-\mu) \le \frac{Z^*_1(t)}{P(t)} \le (1+\lambda)$).

Consider for instance $Z^*_1(0) = (1+\lambda)S_0$ and $\theta^*_1(.) = \sigma(.)$. In that case, $(1, Z^*_1(.)) \in D$ if and only if:

$$0 \le \int_0^t [\mu(s) - r(s)]\, ds \le \ln\left[\frac{1+\lambda}{1-\mu}\right].$$

This condition is satisfied if

$$r(.) \le \mu(.) \le r(.) + \rho \text{ for some } 0 \le \rho \le \frac{1}{T}\ln\left[\frac{1+\lambda}{1-\mu}\right].$$

Also:

$$V^*_{B,T^+} = V^*_{B,T} + f(V^*_{S,T}) = J\left(\frac{\zeta^*}{B(T)}\right) = V_{B,0}B(T). \tag{8.79}$$

The no-trading strategy $(L^* \equiv 0, M^* \equiv 0)$ is optimal and leads to

$$V^*_{B,T} = V_{B,0}B(T) \text{ and } V^*_{S,T} = 0. \tag{8.80}$$

Note that if $\mu(.) \equiv r(.)$, even with no transaction cost, it is not optimal to trade. However, if $\mu(.) > r(.)$, the optimal stock weight is positive if there is no transaction cost. This is true even with transaction costs for infinite-horizon with constant coefficients, as seen in previous section. $\square$

Example 8.11
Consider the case $\mu(.) \equiv r(.)$ deterministic and a strictly positive amount $V^*_{S,0}$ initially invested on the stock. In that case, we have:

$$Z^*_0(0) \equiv 1 \text{ and } Z^*_1(t) = (1-\mu)S_0 \exp\left[\int_0^t \sigma(u)dW_u - \frac{1}{2}\int_0^t \sigma^2(u)du\right].$$

The optimal strategy is given by:

$$L^*(.) \equiv 0, M^*(.) \equiv V_{S,0}\mathbb{I}_{]0,T]}(.).$$

This corresponds to an immediate liquidation of the stock position. We have:

$$V^*_{B,T^+} = V^*_{B,T} + f(V^*_{S,T}) = J\left(\frac{\zeta^*}{B(T)}\right) = \left(V_{B,0} + V_{S,0}(1-\mu)\right)B(T). \quad (8.81)$$

and

$$V^*_{B,t} = \left(V_{B,0} + V_{S,0}(1-\mu)\mathbb{I}_{]0,T]}(t)\right)B(t),\ V^*_{S,t} = V_{S,0}\mathbb{I}_{]0,T]}(t). \quad (8.82)$$

$\square$

8.4 Other frameworks

8.4.1 Labor income

We examine the portfolio optimization problem of an investor who is endowed with a stochastic insurable stream throughout his lifetime. The investor's wealth is assumed to be non-negative over the lifetime interval, which implies that the wage income cannot be sold in the financial market. This liquidity constraint is examined by Cuoco [134], and by El Karoui and Jeanblanc-Picqué [190], who succeed in finding a closed formula.

- *The financial market:*
This consists of one "riskless" asset and d risky securities.

- The riskless asset B satisfies:

$$dB_t = B_t r_t dt,$$

where r_t is the short interest rate and $B_0 = 1$.

- Let S be the price vector of d risky securities. Let $(\Omega, \mathbb{P})$ be the probability space. Assume that S is defined from the following stochastic differential equation (SDE):

$$dS_t = S_{t\_}.\left(\mu(t, S_t)dt + \sigma(t, S_t)dW_t\right), \tag{8.83}$$

where $W = (W_1, ..., W_n)$ is a standard $d-$multidimensional Brownian motion.

- The information is modelled by the filtration $\mathcal{F}_t$ generated by the Brownian motion (and as usual completed in order to contain all $\mathbb{P}$-null sets).

- The processes $\mu(.,.) = (\mu_1, ..., \mu_d)(.,.)$ and $\sigma(.,.) = [\sigma_{i,j}(.,.)]_{i,j}$ satisfy usual conditions which ensure that the previous SDE has one and only one solution. These processes are assumed to be $\mathcal{F}_t$-predictable and uniformly bounded. The matrix $\sigma(t,.)$ is invertible with bounded inverse for any $t \in [0, T]$.

- There exists a predictable and bounded process θ, the risk premium vector, such that:

$$\sigma_t \theta_t = \mu_t - r_t \mathbb{I}, \text{ a.s.}$$

Under the previous hypothesis, the financial market is complete and without arbitrage opportunity.

- *Assumptions on portfolio strategies:* At any time t, the investor receives an income at the rate e_t and chooses the amount c_t per time unit, which is assigned to his consumption, and also the portfolio weighting w_t which is supposed to be self-financing. The cumulative consumption $\int_0^t c_s ds$ is an $\mathcal{F}_t$-adapted-process with $\int_0^t c_s ds < \infty$, $\mathbb{P}$-a.s.

Note that if we impose $c_t \geq \max(e_t, 0)$, then the optimization problem looks like the standard one, since the assumption of non-negative terminal wealth implies the liquidity constraint. When $e_t \geq 0, \forall t \geq 0$, the process e can be interpreted as a labor wage. When $e_t \leq 0, \forall t \geq 0$, the process e can be viewed as a constraint on the consumption process: the excess consumption, $\widetilde{c}_t = c_t - e_t \geq -e_t$.

The portfolio weighting w_t is predictable and such that $\int_0^t \|w_s\|^2 ds < \infty$, $\mathbb{P}$-a.s., where $\|.\|$ denotes the norm.

Thus, the portfolio valueV_t is an Ito process defined by:

$$V_t = V_0 + \int_0^t r_s V_s ds + \sum_{i=1}^{d} \int_0^t w_{i,t} V_s \frac{dS_{i,s}}{S_{i,s}} - \int_0^t (c_s - e_s)\, ds. \tag{8.84}$$

Assumption on the income process: The income process $(e_t)_t$ is spanned by the market assets. There is no "extra noise" on the income dynamics. This means that $(e_t)_t$ is $\mathcal{F}_t$-adapted.

The process $(e_t)_t$ is also supposed to be square-integrable.

- *Assumption on liquidity constraint:* The wealth process V satisfies $V_t \geq 0$ at any time during the management period $[0, T]$. This condition implies that the investor cannot borrow against future labor income.

- *Assumptions on utility functions:* For any time t, let $U(., t)$ and $\widetilde{U}(., t)$ be two utility functions satisfying $\forall t \in [0, T]$,

- $U(., t)$ and $\widetilde{U}(., t)$ are defined on $\mathbb{R}^+$, strictly concave, non-decreasing and continuously differentiable.

- $\lim_{x\to\infty} \frac{\partial}{\partial x} U(x, t) = 0$ and $\lim_{x\to\infty} \frac{\partial}{\partial x} \widetilde{U}(., t) = 0$.

- $U(., .)$ and $\widetilde{U}(., .)$ are continuous on $\mathbb{R}^+ \times [0, T]$ (for example: $U(x, t) = e^{-\rho t} u(x)$ with ρ positive scalar).

Note that the marginal utilities $\frac{\partial}{\partial x} U(x, t)$ and $\frac{\partial}{\partial x} \widetilde{U}(., t)$ are non-decreasing. Therefore, their inverse functions J and $\widetilde{J}$ exist.

As in previous sections, $\widehat{U}$ and $\widehat{\widetilde{U}}$ denote respectively the convex conjugate functions of U and $\widetilde{U}$.

The maximization of the expected intertemporal utility along the time period $[0, T]$ is the following optimization problem:

$$\max_{c,w} \mathbb{E}\left[\int_0^T U(c_s, s)ds + \widetilde{U}\left(V_T, T\right)\right], \tag{8.85}$$

for a given initial budget (at time $t = 0$, the portfolio value V is equal to a given value V_0) and for a given income process e.

Since the financial market is complete, all positive and square-integrable consumption/wealth plans (c, ξ_T) are replicated by square-integrable processes (w, V) such that:

$$dV_t = r_t V_t dt + \sum_{i=1}^{d} w_t \sigma_t V_t \left(dW_t + \theta_t dt\right) - \left(c_t - e_t\right) dt, \tag{8.86}$$

$$V_T = \xi_T. \tag{8.87}$$

Assume that the consumption/wealth plan (c, ξ_T) is such that c is a non-negative $\mathcal{F}_t$-progressively measurable process with $\int_0^t |c_s|\, ds < \infty$, and ξ_T is a non-negative $\mathcal{F}_T$-measurable random variable. Then consider the minimal equation of 8.86 which is:

$$V_t = \mathbb{E}\left[\int_t^T Z_s^t (c_s - e_s) ds + Z_T^t \,|\mathcal{F}_t\right],$$

where $(Z_s^t)_{s \geq t}$ is the shadow state-price process defined by the following forward equation:

$$dZ_s^t = -Z_s^t \left(r_s ds + \theta_s dWs\right), s \geq t,\ Z_t^t = 1. \tag{8.88}$$

Thus, the liquidity constraint has the following form:

$$V_t \geq -I_t, t \geq 0, \tag{8.89}$$

where $I_t = \mathbb{E}\left[\int_t^T Z_s^t e_s ds \,|\mathcal{F}_t\right]$ is the unique solution of 8.86, when $c = e = 0$.

The existence of an optimal solution can be proved in a general setting, as shown in El Karoui and Jeanblanc-Picqué [190] (see Theorems (3.3), (3.7), and (3.8)), by using the Kuhn-Tucker multiplier method to linearize the budget constraint, and by solving a stochastic control problem which is the dual of the unconstrained problem.

Consider the standard Markovian framework: the state-price density Z and the income process e are Markov processes with

$$dZ_t = -Z_t \left[r dt + \theta dW_t\right] \text{ and } de_t = -e_t \left[\mu_e dt + \sigma_e dW_t\right], \tag{8.90}$$

where r, θ, μ_e and σ_e are constant coefficients.

The free dual value function $\mathcal{J}$ is defined by:

$$\mathcal{J}(t, \nu, \varepsilon) = \mathbb{E}\left[\int_t^T \left(\widehat{U}\left[\nu Z_s^t, s)\right] + \nu Z_s^t e_s\right) ds + \widehat{\widehat{U}}\left(\nu Z_T^t\right) |e_t = \varepsilon\right], \tag{8.91}$$

with

$$\widehat{U}\left[\nu, s\right] + \nu\varepsilon + \mathcal{L}(\mathcal{J}) = 0, \tag{8.92}$$

where

$$\mathcal{L}(\mathcal{J}) = \frac{\partial \mathcal{J}}{\partial t} - \nu r \frac{\partial \mathcal{J}}{\partial \nu} - \varepsilon \mu_e \frac{\partial \mathcal{J}}{\partial \varepsilon} + \frac{1}{2}\nu^2\theta^2 \frac{\partial^2 \mathcal{J}}{\partial \nu^2} + \frac{1}{2}\varepsilon^2\sigma^2 \frac{\partial^2 \mathcal{J}}{\partial \varepsilon^2} + \nu\theta\varepsilon\sigma_e \frac{\partial^2 \mathcal{J}}{\partial \varepsilon \partial \nu}, \tag{8.93}$$

and the terminal condition $\mathcal{J}(T, \nu, \varepsilon) = \widehat{\widehat{U}}(\nu)$.

The dual value function $\Phi(t, \nu, e_t)$ is defined by:

$$\Phi(t, \nu, e_t) =$$

$$\min_{D \in \mathcal{D}} \mathbb{E}\left[\int_t^T \left(\widehat{U}\left[\nu Z_s^t D_s, s\right)\right] + \nu Z_s^t D_s e_s\right) ds + \widehat{\widehat{U}}\left(\nu Z_T^t D_{T-}\right) | e_t = \varepsilon\right], \quad (8.94)$$

where $\mathcal{D}$ is the class of adapted, right-continuous, non-increasing processes D such that $D_0 \leq 1$ and $D_T = 0$.

Examine the particular case of absolutely continuous parameters D. Then, the problem can be characterized by using variational inequality:

$$\begin{aligned} \frac{\partial \Phi}{\partial \nu} &\leq 0, \\ \widehat{U}\left[\nu, t)\right] + \nu\varepsilon + \mathcal{L}(\mathcal{J}) &\leq 0, \\ \left(\frac{\partial \Phi}{\partial \nu}\right)\left(\widehat{U}\left[\nu, t)\right] + \nu\varepsilon + \mathcal{L}(\mathcal{J})\right) &= 0, \\ \Phi(T, \nu, \varepsilon) &= 0, \end{aligned}$$

which can be written as

$$\max\left(\frac{\partial \Phi}{\partial \nu}, \widehat{U}\left[\nu, t)\right] + \nu\varepsilon + \mathcal{L}(\mathcal{J})\right) = 0. \quad (8.95)$$

Consider the boundary between the set $\left\{(t, \nu, \varepsilon) \,|\, \frac{\partial \Phi}{\partial \nu}(t, \nu, \varepsilon) = 0\right\}$ and the continuation region $\left\{(t, \nu, \varepsilon) \,|\, \frac{\partial \Phi}{\partial \nu}(t, \nu, \varepsilon) < 0\right\}$.

The variational equation is associated with an American put:

$$\Psi(t, \nu, \varepsilon) = \sup_{D \in \mathcal{D}} \mathbb{E}\left[\int_t^T Z_s^t\left(-\widehat{U}'\left[\nu Z_s^t, s)\right] - e_s\right) ds - Z_\tau^t \mathbb{I}_{\tau = T} \widehat{\widehat{U}}'\left(\nu Z_T^t\right) | e_t = \varepsilon\right].$$

Under the risk-neutral probability $\mathbb{Q}$, the process $\overline{W} = W_t + \theta_t$ is a Brownian motion and we have:

$$dZ_t = Z_t\left[\left(-r + \theta^2\right) dt - \theta d\overline{W}_t\right] \text{ and } de_t = -e_t\left[\left(\mu_e - \sigma_e\theta\right) dt + \sigma_e d\overline{W}_t\right]. \quad (8.96)$$

Also,

$$\Psi(t, \nu, \varepsilon) = \sup_{\tau \geq t} \mathbb{E}_{\mathbb{Q}}\left[\int_t^T \left(-\widehat{U}'\left[\nu Z_s^t, s)\right] - e_s\right) ds - \mathbb{I}_{\tau = T} \widehat{\widehat{U}}'\left(\nu Z_T^t\right) | e_t = \varepsilon\right],$$

is solution of

$$\max\left(\Lambda\Psi(t, \nu) - \Psi(t, \nu)\right) = 0, \quad (8.97)$$

where

$$\Lambda\Psi = \frac{\partial\Psi}{\partial t} + (\theta - r)\frac{\partial\Psi}{\partial\nu} + \varepsilon\left(-\mu_e - \theta\sigma_e\right)\frac{\partial\Psi}{\partial\varepsilon} + \frac{1}{2}\nu^2\theta^2\frac{\partial^2\Psi}{\partial\nu^2}$$
$$+\frac{1}{2}\varepsilon^2\sigma_e^2\frac{\partial^2\Psi}{\partial\varepsilon^2} + \nu\theta\varepsilon\sigma_e\frac{\partial^2\Psi}{\partial\varepsilon\partial\nu} - r\Psi - \widehat{U}' - \varepsilon.$$

The optimal wealth is $V_t^* = \mathcal{H}(t, e_t, Z_t^*)$, where

$$\mathcal{H}(t,\varepsilon,\nu) = \mathbb{E}\left[\int_t^T Z_s^t\left(-\widehat{U}'\left[Z_s^{t*}(\nu), s\right)\right] - e_s\right)ds - Z_T^t\widehat{\widehat{U}}'\left(Z_{T-}^{t*}(\nu)\right)|e_t = \varepsilon\right].$$

Using Ito's formula, we deduce the optimal portfolio w^* which is the solution of:

$$\left[-Z_t^*\theta\frac{\partial\mathcal{H}}{\partial\nu} - \varepsilon\sigma_e\frac{\partial\mathcal{H}}{\partial\varepsilon}\right](t,\varepsilon,Z_t^*) = w_t.$$

Example 8.12

A closed formula can be given for the optimal consumption plan in the infinite horizon case. Assume that the utility function $U(c,t)$ is HARA:

$$U(c,t) = e^{-\delta t}\frac{c^{1-\alpha}}{1-\alpha}, \alpha \neq 1, \text{ then } -\widehat{U}'(\nu,t) = \left(\nu e^{\delta t}\right)^{-\frac{1}{\alpha}}.$$

Suppose that:

- A1: The certainty equivalent present value I_0 of the lifetime labor income satisfies:

$$I_0(\varepsilon) = \mathbb{E}\left[\int_0^\infty Z_t e_t dt\right] < \infty,$$

which means that $r+\ \mu_e - \sigma_e\theta > 0$. Then, we have: $I_0(\varepsilon) = B\varepsilon < \infty$ with $B = \left[r + \mu_e - \sigma_e\theta\right]^{-1}$.

-A2: The free dual problem is well-posed:

$$0 < A = \mathbb{E}\left[\int_0^\infty Z_t\left(\nu e^{\delta t}Z_t\right)^{-\frac{1}{\alpha}}dt\right] < \infty.$$

Note that

$$A = \left[\frac{\delta}{\alpha} + \left(\frac{\alpha-1}{\alpha}\right)\left(r + \frac{\theta^2}{2\alpha}\right)\right]^{-1}.$$

<u>For the free case</u>, from Markovian properties, we deduce that the optimal free wealth process V^f is given by:

$$V_t^f(\nu) = A\left[\nu e^{\delta t}Z_t\right]^{-\frac{1}{\alpha}} - Be_t,$$

with the multiplier ν such that the initial wealth is equal to V_0:

$$A\left[\nu\right]^{-\frac{1}{\alpha}} = V_0 + B\varepsilon.$$

We deduce that the optimal consumption c^* is given by:

$$c_t^*(\nu) = \frac{V_t^f(\nu) + Be_t}{A}.$$

The optimal wealth process V^f is the solution of:

$$\begin{aligned} dV_t^f &= -Bde_t + Ad\left(\nu Z_t e^{\delta t}\right)^{-\frac{1}{\alpha}} \\ &= Be_t\left(\mu_e dt + \sigma_e dW_t\right) \\ &\quad + \left(V_t^f + Be_t\right)\left[\left(\frac{1}{\alpha}(r-\delta) + \frac{1}{2\alpha}\left(\frac{\alpha+1}{\alpha}\theta^2\right)\right)dt + \frac{\theta}{\alpha}dW_t\right]. \end{aligned}$$

The optimal portfolio is given by:

$$w_t^* = \frac{\theta}{\alpha}V_t^f(\nu) + \left(\sigma_e + \frac{\theta}{\alpha}\right)Be_t.$$

For the constrained case, the wealth value $V_t(\nu)$ is deduced from the price of an American option written on the negative part of the free value:

$$V_0(\nu) = V_0{}^f(\nu) + \mathcal{J}_0(\nu),$$

where

$$\mathcal{J}_0(\nu) = \sup_\tau \mathbb{E}\left[Z_\tau\left(A\left(\nu e^{\delta\tau}Z_\tau\right)^{-\frac{1}{\alpha}} - Be_\tau\right)\right].$$

The associated one-dimensional stopping time problem is as follows. The value function $\mathcal{J}_0$ corresponds to an optimal stopping problem with a two dimensional Markov diffusion processes $(\nu e^{\delta t}Z_t, e_t, t \geq 0)$. Since the payoff is homogeneous, this problem can be transformed into a one-dimensional stopping time problem:

$$\frac{\mathcal{J}_0(\nu)}{\varepsilon} = \sup_\tau \mathbb{E}\left[\frac{Z_\tau e_\tau}{\varepsilon}\left(AY_\tau - B\right)^-\right] \text{ with } Y_t = \frac{\left(\nu e^{\delta t}Z_t\right)^{-\frac{1}{\alpha}}}{e_t}.$$

The function $\frac{\mathcal{J}_0(\nu)}{\varepsilon}$ is the value function of an optimal stopping time problem w.r.t. $(AY - B)$, where the Markov process $(Y_t)_t$ is a geometrical Brownian motion under the risk-neutral probability $\mathbb{Q}^e$ with characteristics given by:

$$\begin{cases} dY_t = Y_t\left(\mu_Y + \sigma_Y dW_t^Y\right), Y_0 = \nu^{-\frac{1}{\alpha}}\varepsilon^{-1}, \\ \mu_Y = B^{-1} - A^{-1}, \sigma_Y^2 = \frac{\theta^2}{\alpha^2} + \sigma_e^2 + \frac{\theta\sigma_e}{\alpha}. \end{cases}$$

The optimal stopping problem is solved as follows. Since the value function $\frac{\mathcal{J}_0(\nu)}{\varepsilon}$ is convex and monotonous, the optimal stopping problem is equivalent to the search of an entrance time $\tau(a)$ such that

$$\tau(a) = \inf\{t : Y_t \leq a\}.$$

To this stopping time is associated a "reward" given by:

$$\Psi(y,a) = \mathbb{E}_{\mathbb{Q}^e}\left[e^{-\tau(a)/B}\left(AY_{\tau(a)} - B\right)^-\right] = (A\min(a,y) - B)^- \,\mathbb{E}_{\mathbb{Q}^e}\left[e^{-\tau(a)/B}\right].$$

We can use the Laplace transform of the law of a Brownian entrance time to compute $\Psi(y,a)$. Let $T(b,\mu) = \inf\{t : \mu t + W_t = b\}$ be the entrance time for the Brownian motion with drift $(\mu t + W_t)_t$:

$$\mathbb{E}\left[e^{-\lambda T(b,\mu)}\mathbb{I}_{T(b,\mu)<\infty}\right] = \exp\left[b\mu - |b|\sqrt{\mu^2 + 2\lambda}\right], \lambda > 0.$$

The process $\ln(Y)$ is a Brownian motion with drift ∇ given by:

$$\nabla = \mu_Y - \frac{1}{2}\sigma_Y^2 = \frac{r - \delta - \theta^2/2}{\alpha} + \mu_e - \frac{\sigma_e^2}{2} - \theta\sigma_e - \frac{\theta\sigma_e}{2\alpha}.$$

Thus, the stopping time $\tau(a)$ corresponds to an entrance time for a Brownian motion:

$$\tau(a) = T\left(\frac{1}{\sigma_Y}\ln\left(\frac{a}{y}\right), \frac{\nabla}{\sigma_Y}\right), \text{ for } a < y.$$

We have

$$\mathbb{E}_{\mathbb{Q}}\left[e^{-\tau(a)/B}\right] = \left(\frac{\min(a,y)}{y}\right)^{\Delta},$$

with

$$\Delta = \frac{1}{\sigma_Y^2}\left[\nabla + \sqrt{\nabla^2 + 2\sigma_Y^2 B^{-1}}\right].$$

The optimal stopping time corresponds to the value of a which maximizes $\Psi(y,a)$:

$$a^* = \frac{\Delta B}{A(1+\Delta)} \text{ (note that: } a^* \leq 1).$$

We have also:

$$\frac{\mathcal{J}_0(\nu)}{\varepsilon} = \Psi(y, \min(a^*, y)) = (B - A\min(a^*, y))^+ \left(\frac{\min(a^*, y)}{y}\right)^{\Delta}.$$

El Karoui and Jeanblanc [190] prove that there is a closed form solution for the optimal wealth associated with the constrained dual problem. It is based

on the free boundary of this problem:

$$
\begin{aligned}
b(\varepsilon) &= \left(\frac{A\,(1+\Delta)}{\Delta B\varepsilon}\right)^{\alpha}, \\
Z_0(\nu) &= 0 \text{ if } \nu \geq b(\varepsilon), \\
Z_0(\nu) &= \frac{B\varepsilon}{1+\Delta}\left(\frac{\Delta B\varepsilon\nu^{\frac{1}{\alpha}}}{A\,(1+\Delta)}\right)^{\Delta} + A\nu^{-\frac{1}{\alpha}} - B\varepsilon, \text{ otherwise.}
\end{aligned}
$$

Finally, we deduce that the optimal consumption function is given by: $(x = V_0)$

$$c^*(\nu_x) = \min\left(\nu_x, b(\varepsilon)\right)^{-\frac{1}{\alpha}} = \varepsilon(\max\left(y, a^*\right)) \text{ with } y = \varepsilon^{-1}\nu_x^{-\frac{1}{\alpha}}.$$

The optimal consumption and optimal wealth are linked through a feedback formula:

$$x^* = \varepsilon\left(\frac{B}{1+\Delta}\left(\frac{\Delta B}{A\,(1+\Delta)}\frac{\varepsilon}{c^*}\right)^{\Delta} + A\frac{c^*}{\varepsilon} - B\right). \tag{8.98}$$

When the initial wealth value is equal to 0, the maximal value for the consumption function is a fraction equal to a^* of the income stream.

The fraction a^* is equal to $\frac{\Delta B}{A(1+\Delta)}$, and is strictly smaller than the fraction $\frac{B}{A}$ corresponding to the free problem.

The following figure illustrates how the optimal consumption depends on the optimal wealth, according to relation (8.98). It is drawn in the plane $(\frac{x^*}{\varepsilon}, \frac{c^*}{\varepsilon})$. The upper curve corresponds to the solution of the free problem.

The parameter values are:

$r = 0.1$, $b = 0.2$, $\sigma = 0.1$, $\mu_e = 1$, $\sigma_e = 0.1$, $\gamma = 0.7$ and $\beta = 1$.

$\square$

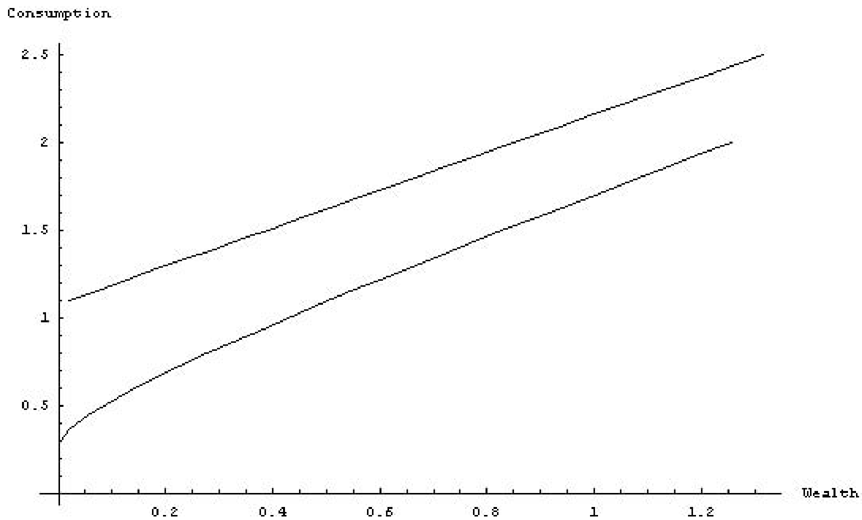


FIGURE 8.2: Optimal consumption as function of optimal wealth

8.4.2 Stochastic horizon

If the investor does not know with certainty the time of exiting the market (*e.g.*, retirement, death, *etc.*), the time horizon becomes uncertain. Therefore, we have to examine optimization problem such as:

$$\max \mathbb{E}\left[U(V_{\tau\wedge T})\right], \tag{8.99}$$

where $\tau \wedge T$ is the investment horizon.

When τ is a stopping time and the financial market is complete, the solution is provided, for example, by Karatzas and Wang [322] and Richard [424]. Blanchet-Scalliet *et al.* [79] consider the case where the time τ is a random time horizon and the conditional distribution of τ given the available information at time t is known. The process $F_t = \mathbb{P}\left[\tau \leq t \,|\mathcal{F}_t\right]$ satisfies the so-called (G)-assumption. The process $(F_t)_t$ is non-decreasing and right-continuous. In the complete market framework, they examine in particular the case where F_t admits a density function f_t w.r.t. the Lebesgue measure. Then, problem (8.99) is equivalent to:

$$\max \mathbb{E}\left[\int_0^T U(V_t) f_t dt + U(V_T)\left(1 - \int_0^T f_t dt\right)\right]. \tag{8.100}$$

This latter problem is reduced to a standard stochastic control problem for which we can use dynamic programming. Using a mild assumption on the density f, this problem is solved through PDE (see Blanchet-Scalliet et al.

[79]) or BSDE (see El Karoui et al.[191]). However, they do not include either the case where $F_t = \mathbb{I}_{[\tau \leq t]}$, nor the case where F has no density w.r.t. the Lebesgue measure.

This the reason why Bouchard and Pham [84] introduced a more general *wealth path dependent* utility maximization problem:

$$\max \mathbb{E}\left[\int_0^T U(V_t)dF_t\right], \tag{8.101}$$

where the process $(F_t)_t$ is non-decreasing and right-continuous. Note that

$$F_t = \mathbb{P}\left[\tau \wedge T \leq t \,|\mathcal{F}_t\right].$$

Note also that, contrary to fixed time horizon, the whole path of the portfolio process must be taken into account. Therefore, the dual variables are stochastic processes and not simple $\mathcal{F}_T-$measurable random variables. Thus, the optimization problem is not reduced to a static one. Bouchard and Pham [84] derive a dual formulation in a general incomplete semimartingale framework. Then, they provide a solution to the primal problem, using a calculus of variation on the primal problem. The model is defined as follows:

- *The financial market* consists of one bond B chosen as numeriare and d securities $(S_{i,t})_{i,t}$. The process S is assumed to be a semimartingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$.

- *Portfolio strategy and value:* The numbers of shares $(\theta_{i,t})_{i,t}$, invested on each asset i at time t, is a predictable process, which is integrable w.r.t. the price process S. The portfolio value process $(V_t)_t$, associated to strategy $(\theta_{i,t})_{i,t}$, is given by, for $t \in [0, T]$,

$$V_t = V_0 + \sum_{i=1}^{d} \int_0^t \theta_{i,s} dS_{i,s}. \tag{8.102}$$

The utility function U satisfies the same assumptions as in previous sections.

The optimization is well-defined by assuming that the wealth process V is in the set

$$\mathcal{A} = \left\{V : \forall t \in [0, T], V_t \geq 0, a.s. \text{ and } \mathbb{E}\left[\int_0^T U^-(V_t)dF_t\right] < \infty\right\}.$$

To avoid degeneracy, it is assumed that $\mathbb{P}\left[F_T > 0\right] > 0$. The random variable F_T is also supposed to satisfy $\mathbb{E}\left[F_T\right] < \infty$ and, without loss of generality, $\mathbb{E}\left[\int_0^T dF_t\right] = 1$.

PROPOSITION 8.5 *Bouchard and Pham [84]*
Under the previous hypothesis and assuming also that:
i) The financial market has no arbitrage opportunity (there exists at least one equivalent local martingale)
ii) The value function $\widehat{\mathcal{J}}$ *of the problem*

$$\widehat{\mathcal{J}}(y) = \inf_{Y \in \mathcal{D}(y)} \int_0^T \widehat{U}(Y_t) dF_t < \infty,$$

where $\mathcal{D}(y)$ *is the set of all processes* Y *such that* $Y \geq 0$ *w.r.t. the measure* dm*, which is defined by:*

$$m(E \times F) = \mathbb{E}\left[\int_0^T \mathbb{I}_{E \times F} dF_t\right].$$

Then, we have:
i) Existence of a solution to the primal problem: for all $V_0 > 0$*, there exists a unique optimal solution* V^* *in the set* $\mathcal{A}$*.*
ii) Existence of a solution to the dual problem: for all $Y_0 > 0$*, there exists a unique optimal solution* Y^* *in the set* $\mathcal{D}(Y_0)$*.*
iii) Duality relations: (notation: $J = (U')^{-1}$*) for all* $V_0 > 0$*,*

$$V_t^* = J(Y_t^*) \text{ with } V_0 Y_0^* = \mathbb{E}\left[\int_0^T J(Y_t^*) Y_t^* dF_t\right].$$

Example 8.13 *Blanchet-Scalliet et al. [79]*
The financial market contains a riskless asset with rate r and a risky asset S which is assumed to be a semimartingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ solution of the SDE:

$$dS_t = S_{t^-}. \left[\mu dt + \sigma dW_t\right], \tag{8.103}$$

where W is a standard Brownian motion and μ is constant and σ is a non-negative constant. This financial market is complete. Thus, there exists one and only one risk-neutral probability $\mathbb{Q}$ defined with the relative risk process η by $\eta = \sigma^{-1}\left[\mu - r\mathbb{I}\right]$. Assume that the deterministic time horizon is infinite ($T < +\infty$). Introduce:

$$\mathcal{J}(t, V_0) = \sup_w \mathbb{E}\left[\int_t^\infty f(u) U(V_u) du\right].$$

Case 1: For some deterministic function f, $F(t) = \int_0^t f(u) du$ on $[0, T]$ and $F(T) = 1$. The random time τ is independent from the filtration and has the density f.

- If $U(x) = \ln x$, we have:

$$\mathcal{J}(t, V_0) = p(t) + q(t) \ln V_0,$$
$$w_t^* = \frac{\mu - r}{\sigma^2} V_t^*,$$

with

$$q(t) = 1 - \int_0^t f(u)du, \ p(t) = \lambda - ct + c \int_0^t \int_0^s f(u)du,$$
$$c = r + \frac{1}{2} \frac{(\mu - r)^2}{\sigma^2}, \ \lambda = \mathbb{E}[\tau] \left[r + \frac{1}{2} \frac{(\mu - r)^2}{\sigma^2} \right].$$

- If $U(x) = \frac{x^\alpha}{\alpha}, 0 < \alpha < 1$, we have:

$$\mathcal{J}(t, V_0) = \widetilde{p}(t) V_0^\alpha,$$
$$w_t^* = -\frac{\mu - r}{(\alpha - 1)\sigma^2} V_t^*,$$

with

$$\widetilde{p}(t) = e^{-at} \int_t^\infty e^{au} f(u)du, \ \widetilde{c} = r - \frac{1}{2(\alpha - 1)} \frac{(\mu - r)^2}{\sigma^2}.$$

- If $U(x) = e^{-x}, r = 0$ and $\mathbb{E}\left[e^{\left(\frac{(\mu - r)^2}{\sigma^2} \right) \tau} \right] < \infty$, then we have:

$$\mathcal{J}(t, V_0) = \widetilde{p}(t) e^{-V_0},$$
$$w_t^* = \frac{\mu}{\sigma^2},$$

with

$$\widetilde{p}(t) = e^{-\widehat{c}t} \int_t^\infty e^{\widehat{c}u} f(u)du, \widehat{c} = \frac{1}{2} \frac{(\mu - r)^2}{\sigma^2}.$$

<u>Case 2:</u> $F(t) = \mathbb{P}[\tau \leq t \,|\mathcal{F}t] = \int_0^t f(u)du$ on $[0, T]$ where f is solution of the following SDE:

$$df_t = f_t (adt + bdW_t).$$

The random time τ is a stopping time w.r.t. the filtration.

- If $U(x) = \ln x$ and $a < 0$, we have:

$$\mathcal{J}(t, V_0, f_0) = f_0 \left[C_2 e^{-C_1 t} \int_t^T q(s) e^{C_1 s} ds + q(t) \ln V_0 \right],$$
$$w_t^* = \left(\frac{\mu - r + \sigma b}{\sigma^2} \right) V_t^*,$$

with

$$q(t) = \left(\frac{1}{a} + \frac{1}{C_1}\right) e^{C_1(T-t)} - \frac{1}{C_1}, C_1 = b\left(\frac{\mu - r}{\sigma} + b\right),$$

$$C_2 = r + (\mu - r)\left(\frac{\mu - r + \sigma b}{\sigma^2}\right) - \frac{1}{2}\left(\frac{\mu - r + \sigma b}{\sigma}\right)^2.$$

- If $U(x) = \frac{x^\alpha}{\alpha}$ and $a < 0$, we have:

$$\mathcal{J}(t, V_0) = f_0 \frac{V_0^\alpha}{\alpha}\left(\frac{1}{a} + \frac{1}{C} e^{CT}\right) e^{-Ct} - \frac{1}{C},$$

$$w_t^* = -\frac{\mu - r + \sigma b}{(\alpha - 1)\sigma^2} V_t^*,$$

with $C = \alpha\left(r + (\mu - r)\left(\frac{\mu - r + \sigma b}{(\alpha-1)\sigma^2}\right)\right) + \frac{1}{2}\alpha(\alpha - 1)\left(\frac{\mu - r + \sigma b}{(\alpha - 1)\sigma}\right)^2 - b\left(\frac{\mu - r + \sigma b}{(\alpha - 1)\sigma}\right)$.

$\square$

8.5 Further reading

Standard results about dynamic portfolio optimization can be found in Korn [333]. As shown in Schachermayer [452], the asymptotic elasticity condition is related to the asymptotic relative risk aversion under mild assumptions.

Investment-consumption models with transaction costs are examined in Akian *et al.* ([12],[13]). Quadratic optimization with transaction costs is analyzed by Adcock and Meade [6]. Bielecki and Pliska [70] study risk-sensitive dynamic asset management with transaction costs. Optimal portfolio management with transaction costs equal to a fixed fraction of the portfolio, corresponding to a portfolio management fee, is examined in Morton and Pliska [394]; the asymptotic growth rate (the "Kelly criterion") is maximized on an infinite-horizon. It is shown that even with very low transaction costs, the optimal solution leads to very infrequent trading. Duffie and Sun [177] examine a model where the transaction is also a fixed fraction of portfolio value and also with a proportional cost for withdrawal of funds for consumption. Assuming that the wealth is only observed at transaction times, it is optimal to trade at fixed deterministic intervals. Fleming and Soner [232] also provide details about viscosity solutions to portfolio optimization with transaction costs (see also Lions [358] and Zariphopoulou [510]). Cadenillas *et al.* ([97],[98]) also study portfolio optimization with transaction costs and possible taxes. Portfolio optimization with transaction costs based on impulse control is introduced in Korn [334]. Deelstra *et al.*[153] provide the dual formulation of the utility maximization under transaction costs.

Portfolio optimization with labor income has been also studied without the assumption that wealth is non-negative over the period of trading, by Karatzas *et al.* [317]. The investor is allowed to capitalize his wage income at some interest rate which is simply added to the current wealth. He and Pages [287] consider that the investor's wealth must be non-negative. They prove that there exists an optimal solution by application of duality theory. The optimization problem is transformed in an unconstrained dual shadow price problem. Assuming that the asset price is a Markov-diffusion process, they solve the problem by the dynamic programming method when the labor income is a function of the asset price. Duffie *et al.* [174] examine the same kind of optimization problem by dynamic programming when the income process is uninsurable. Thus, the financial market is incomplete since the labor income cannot be duplicated by a portfolio. They provide a quasi-explicit solution of the H.J.B. equation when the utility function is HARA, and when coefficients are deterministic. They show that at a zero wealth, the investor consumes a fixed fraction of wealth while saving the remaining amount in the riskless asset.

Portfolio optimization under partial information has been studied by Lakner [343]. Nagai and Peng [397] consider an optimal investment problem with partial information for a factor model. Jeanblanc *et al.* [296] examine utility maximization under partial information when asset dynamics are modelled by a jump-diffusion process. They assume that only the vector of stock price is observable. However, they prove that the optimization problem can be rewritten as a problem with coefficients depending on past history of observed prices.

Portfolio optimization with possible bankruptcy can also be analyzed. For example, the investor is under the obligation to pay a debt until he declares bankruptcy. This type of problem has been examined by Cadenillas and Sethi [99], Lehoczky *et al.* [350], Sethi [460] and Jeanblanc and Lakner [297].

Dynamic benchmark optimization is analyzed by Browne [94], who searches to outperform a stochastic benchmark. Buckley and Korn [96] use impulse control to determine an optimal index tracking under transaction costs.

Part IV

Structured portfolio management

"The sophistication of portfolio insurance users has grown as rapidly as the product itself... Sponsors are realizing the advantage of programs that protect a fund's surplus. Portfolio insurance is being applied to many different classes of assets besides equities; fixed-income and international investments are growing areas of application. An early criticism of portfolio insurance was that it reduced return as well as reducing risk. But users are discovering that portfolio insurance can be used aggressively rather than simply to reduce risks. Long-run returns can actually be raised, with downside risks controlled, when insurance programs are applied to more aggressive active assets. Pension, endowment, and educational funds can actually enhance their expected returns by increasing their commitment to equities and other high-return sectors, while fulfilling their fiduciary responsibilities by insuring this more aggressive portfolio. Compared with current static allocation techniques, annual expected returns can be raised by as much as 200 basis points per year... Dynamics can be used to mold a set of returns to virtually any feasible investor objective. It can be used to manage the risks of corporate balance sheets as well as investment funds. As such, it may represent the most significant advance to date in the science of financial engineering."

Hayne E. Leland and Mark Rubinstein, "The Evolution of Portfolio Insurance," published in *Dynamic Hedging: A Guide to Portfolio Insurance*, (1988).

Chapter 9

Portfolio insurance

The purpose of portfolio insurance is to limit portfolio losses when sudden drops occur in the financial market, while allowing investors to benefit from potential rises of the market. Thus, very often, for insured portfolio values at maturity:

- There exists a guaranteed amount (it means that the probability that the guarantee is violated must be equal to 0).
- When the market rises, the portfolio return must also rise at (at least) a predetermined percentage of a given index return.

Therefore, portfolio insurance generally requires to specify the guarantee constraint and the portfolio maturity.

Two standard portfolio insurance methods are the *Option Based Portfolio Insurance* (OBPI) and the *Constant Proportion Portfolio Insurance* (CPPI):

- The OBPI, introduced by Leland and Rubinstein [349], consists of a portfolio invested in a risky asset S (usually a financial index such as the $S\&P$) covered by a listed put written on it.
- The CPPI was introduced by Perold [406] (see also Perold and Sharpe [407]) for fixed-income instruments and Black and Jones [76] for equity instruments. This method uses a simplified strategy to allocate assets dynamically over time.

This chapter provides:

- First, some basic properties of OBPI and CPPI methods, such as their payoff functions, for instance in the geometric Brownian motion framework.
- Second, their payoffs at maturity are compared by means of stochastic dominance, by means of four first moments of their returns and of the cumulative distribution of their ratio.
- Finally their dynamics properties are examined and compared. It is proved that the OBPI method is a generalized CPPI where the multiple is allowed to vary. We will also focus on the dynamics of both methods, in particular their "Greeks." In what follows, we use in particular results in Bertrand and Prigent ([60], [58], [59]).

9.1 The Option Based Portfolio Insurance

Leland and Rubinstein [349] introduce the OBPI method which uses combinations of standard securities (such as bonds and stocks) and options.

To illustrate this method, let us examine the following example.

Example 9.1 **Why do we need options?**
Consider a given time horizon T (for example one year).

- A first investor, denoted by I_1, chooses to invest in a riskless asset B and in a risky security S (typically a financial index).

- A second investor, denoted by I_2, chooses to invest in B and to buy a Call option C, written on the asset S.

The amount in B of investor I_2 corresponds to the discounted exercise price of the option.

Assume that:

- Both investors have the same initial total wealth V_0 invested in the financial market.

- Each of them wants to recover at maturity a same given percentage p of his initial investment.

Suppose that the riskless rate r is equal to 3%, and that the volatility of S is equal to 20%. Assume also that $B_0 = 1$ and $S_0 = 100$. Within this framework, the value C_0 of the at-the-money Call option is equal to $C_0 = 9.41$.

- The initial value V_0 of investor I_2, who buys the option, is given by:

$$V_0 = S_0 e^{-rT} + C_0 \simeq 106.41.$$

Then, the portfolio value $V_T^{(2)}$ of investor I_2 at maturity T is equal to:

$$V_T^{(2)} = S_0 + (S_T - S_0)^+ = 100 + (S_T - 100)^+.$$

This value is higher than the amount $S_0 = 100$, which corresponds to the guaranteed amount for investor I_2. It is the minimal amount that he recovers at maturity if he cannot exercise the option (if the price of S decreases). Note that the guaranteed percentage p is given by $p = S_0/V_0 \simeq 94\%$.

- The initial value V_0 of investor I_1, who chooses a simpler strategy: buy and hold quantities a and b respectively in B and in S, without using an option, is given by

$$V_0 = aB_0 + bS_0 \simeq 106.41.$$

Since investor I_1 wants the same guaranteed percentage as I_2, he must invest the same amount in the riskless asset B. This implies that $aB_0 = S_0 e^{-rT} \simeq 97$.

Therefore, he chooses the same percentage of investment in B (here, about 91%).

Thus, the portfolio value $V_T^{(1)}$ of investor I_1 at maturity T is equal to

$$V_T^{(1)} = aB_0 e^{rT} + bS_T \simeq 100 + 0.094 \times \ S_T.$$

- Let us compare the two portfolio returns for the two cases:

* First case: the growth of the asset price S is equal to 30%.
* Second case: the asset price S decreases by -30%.

For a rise of 30%, the portfolio return of investor I_1 is approximately equal to $p + (0.09) \times 1.3$, which gives a rate of about 5.7%. For a decrease of -30%, the portfolio return of investor I_1 is approximately equal to $p + (0.09) \times 0.7$, which gives a rate of about 0.3%.

For a rise of 30%, the portfolio return of investor I_2 is approximately equal to $p + (0.09)\times$ (option return), which gives a rate of about 22.7%. For a decrease of -30%, the portfolio return of investor I_2 is approximately equal to p, which gives a rate of about -6%.

Therefore, when the risky asset price S increases significantly, the portfolio which contains the option has a much better return than the simple combination of the two basic assets B and S.

However, it is the converse when the risky asset drops. Nevertheless, in that case, the guarantee is always satisfied. Note also that, with respect to a riskless investment with return $e^{rT} \simeq 3\%$, both portfolios clearly provide smaller returns when the financial market is bearish, but a riskless investment cannot benefit from bullish market.

Note that the percentage that the portfolio of the first investor can acquire from a potential increase of the risky asset S is rather low. Indeed, we have:

$$\left(\frac{V_T}{V_0}\right) = p + (1 - pe^{-rT}) \left(\frac{S_T}{S_0}\right) \simeq 0.94 + (0.09) \left(\frac{S_T}{S_0}\right). \tag{9.1}$$

Thus, only about 9 % of the risky return S_T/S_0 can be obtained.

Concerning the portfolio including the option, note that if the option can be exercised, its return is equal to:

$$\left(\frac{V_T}{V_0}\right) = p + (1 - pe^{-rT}) \frac{S_0 \left(\frac{S_T}{S_0} - 1\right)}{C_0} \simeq 0.94 + (0.95) \times \left(\frac{S_T}{S_0} - 1\right). \tag{9.2}$$

Therefore, this portfolio can benefit from a high leverage when the financial market rises.

$\square$

9.1.1 The standard OBPI method

The OBPI, introduced by Leland and Rubinstein [349], consists of a portfolio invested in a risky asset S (usually a financial index such as the S&P) covered by a listed put written on it. Whatever the value of S at the terminal date T, the portfolio value will be always greater than the strike K of the put. At first glance, the goal of the OBPI method is to guarantee a fixed amount only at the terminal date. In fact, as shown in what follows, the OBPI method allows one to get portfolio insurance at any time. Nevertheless, the European put with suitable strike and maturity may be not available on the market. Hence, it must be synthesized by a dynamic replicating portfolio invested in a risk-free asset (for instance, T-bills) and in the risky asset.

The portfolio manager is assumed to invest in two basic assets: a money market account, denoted by B, and a portfolio of traded assets such as a composite index, denoted by S. The period of time considered is $[0, T]$. The strategies are self-financing.

The value of the riskless asset B evolves according to:

$$dB_t = B_t r dt,$$

where r is the deterministic interest rate.

The dynamics of the market value of the risky asset S are given by the standard diffusion process:

$$dS_t = S_t \left[\mu dt + \sigma dW_t\right],$$

where W_t is a standard Brownian motion.

The OBPI method consists basically of purchasing q shares of the asset S and q shares of European put options on S with maturity T and exercise price K.

Thus, the portfolio value V^{OBPI} is given at the terminal date by:

$$V_T^{OBPI} = qS_T + q(K - S_T)^+, \tag{9.3}$$

which is also $V_T^{OBPI} = qK + q(S_T - K)^+$, due to the Put/Call parity. This relation shows that the insured amount at maturity is the exercise price times the quantity q: qK.

The value V_t^{OBPI} of this portfolio at any time t in the period $[0, T]$ is:

$$V_t^{OBPI} = qS_t + qP(t, S_t, K) = qK.e^{-r(T-t)} + qC(t, S_t, K), \tag{9.4}$$

where $P(t, S_t, K)$ and $C(t, S_t, K)$ are the Black-Scholes values of the European put and call.

REMARK 9.1 Assume that standard no-arbitrage conditions hold together with no market friction. Then, for all dates t before T, the portfolio value is always above the deterministic level $qKe^{-r(T-t)}$, which shows that the OBPI strategy also provides a dynamic guarantee. □

REMARK 9.2 1) The amount insured at the final date is often expressed as a percentage p of the initial investment V_0 (with $p \leq e^{rT}$). Since, here, this amount is equal to the strike K itself, it is required that K is an increasing function of the percentage p, determined from the relation:

$$pV_0(K) = p(qK.e^{-rT} + qC(0, S_0, K)) = qK. \tag{9.5}$$

Indeed, we have

$$\frac{C_0(K)}{K} = \frac{1 - p\ e^{-rT}}{p}, \tag{9.6}$$

where $C_0(K)$ denotes the initial Call option value (for example determined from Black and Scholes formula). Since the ratio $\frac{C_0(K)}{K}$ does not depend on $\frac{S_0}{K}$, we deduce that the higher the guaranteed percentage p, the higher must be the strike K (for a given value S_0).

2) For a given initial portfolio value V_0 and for given values S_0 and $P_0(K)$, the number q of shares is a decreasing function of the exercice price K, since q is determined from

$$q = \frac{V_0}{S_0 + P_0(K)}, \tag{9.7}$$

where $P_0(K)$ denotes the initial Put value.

Finally, note that for a given investment V_0 and a fixed guaranteed percentage p, the option strike K and the number of shares q are entirely determined. □

To simplify the presentation and without loss of generality, we shall assume that q is normalized and set equal to one.

Then:

$$V_T^{OBPI} = S_T + (K - S_T)^+ = K + (S_T - K)^+. \tag{9.8}$$

This function is increasing and convex w.r.t. the risky asset price S_T at maturity. Therefore, it has the typical features of the portfolio payoffs with guarantee constraints. Indeed, such portfolio can benefit from a market rise, *since if the risky asset price is higher than the strike at maturity*, the return is given by:

$$\left(\frac{V_T}{V_0}\right) = \left(\frac{S_T}{S_0}\right) \times \left(\frac{S_0}{S_0 + P_0(K)}\right).$$

Thus, in that case, the percentage which is obtained is equal to $\frac{1}{1+P_0(K)/S_0}$ (for the previous example, this percentage is equal to 94.4%. Thus, for a return S_T/S_0 equal to 30%, the return V_T/V_0 is about 22.7%).

The portfolio profile V_T^{OBPI} is as follows (for $K = 100$):

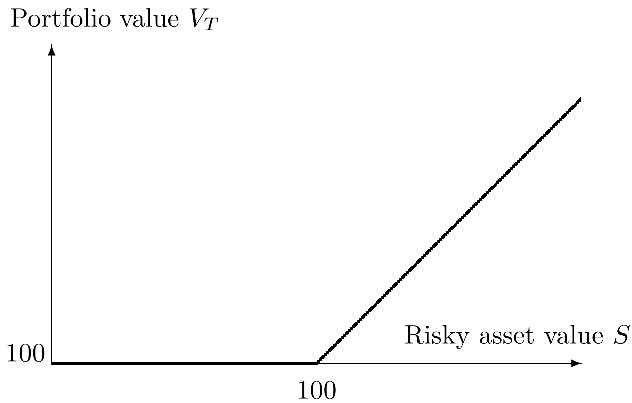


FIGURE 9.1: OBPI portfolio value as function of S

9.1.2 Extensions of the OBPI method

Several extensions can be proposed. They are mainly based on other kinds of options to be included in the portfolio. Among them, more general polynomial options can be substituted for the standard European options. As shown in Chapter 10, the choice of such options can be based on risk aversion and utility maximization.

9.1.2.1 Polynomial options

Assume that the guarantee constraint is as follows. At maturity, the portfolio value V_T must be always higher than

$$F_T = h_c\left(S_T\right).$$

For example, consider a linear function $h_c\left(S_T\right) = aS_T + b$. The investor has a fixed guarantee b whatever the market evolution, and also a minimal percentage a of the potential market rise. Moreover, the underlying asset of the option is a power function of the risky asset price at maturity.

Then, we are led to the following combination:

$$V_T = d.S_T^m + \left([aS_T + b] - d.S_T^m\right)^+ = [aS_T + b] + \left(d.S_T^m - [aS_T + b]\right)^+ . \quad (9.9)$$

For a simple fixed guarantee, we get:

$$V_T = d.S_T^m + \left(K - d.S_T^m\right)^+ = K + \left(d.S_T^m - K\right)^+ . \quad (9.10)$$

The figure 9.2 provides an example of such a power option payoff.

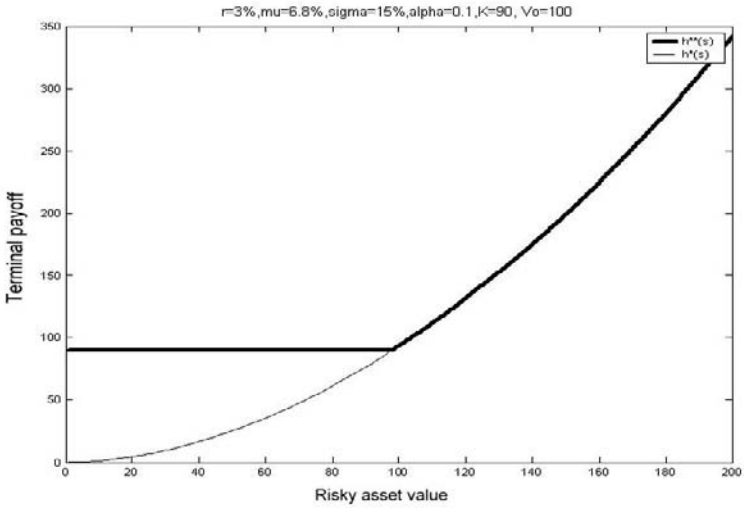


FIGURE 9.2: Call-power option profiles

Figure 9.3 provides examples of paths of call-power portfolios and of standard OBPI one for three types of paths of the risky asset S and three different values of m. The OBPI strategy is obviously a special case of call power option with $m = 1$ (linear case).

In order to compare the two methods, the same path of S (drop, rise and stability) corresponds to each row.

We note that the guaranteed level strongly depends on the value of the parameter m. This result is quite intuitive since, in order to get a terminal payoff more and more convex, the investor must more and more bear the risk of a portfolio value drop.

The first column provides the path of Call-power options with $m = 0.6 < 1$. Above the guaranteed level, the payoff is concave. For a bearish market, the portfolio value reachs the floor very quickly but remains always above the OBPI payoff. For a bullish market, the concavity of the payoff implies that the portfolio value is smaller than the OBPI one. Then, the concavity allows the "reduction" of the variations of the risky asset, as illustrated by the last graph in the first column.

The second column corresponds to a medium case with $m = 2$. The guaranteed level is reduced from 98.96% (previous case) to 83.72% of the initial invested amount, as illustrated by the first graph corresponding to a market decrease. The last graph shows that, contrary to the concave case, the risky

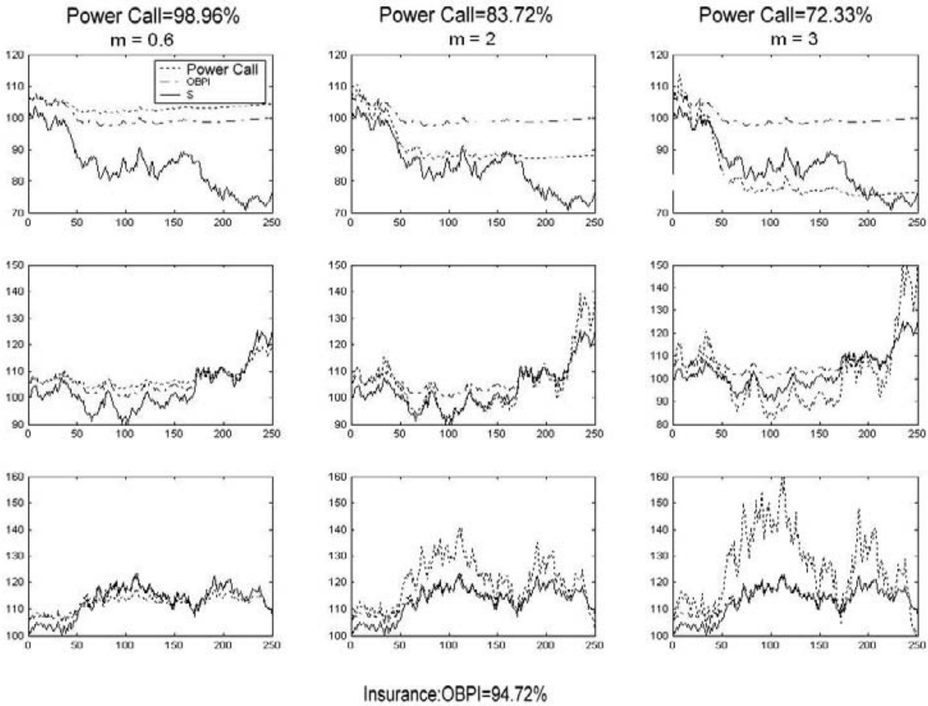


FIGURE 9.3: Call-power option paths

asset variations are amplified.

The last column corresponds to the smaller guaranteed level with very volatile paths. The next table summarizes informations about the four first moments: mean, standard deviation, skewness, and kurtosis, and also the guaranteed levels for the different values of m. Let $V_0 = 100$.

TABLE 9.1: OBPI and Call-power moments

m	Guarantee	$E(V_T)$	$\sigma(V_T)$	Skew	Kurto
0.6 (concave)	98.96%	103.14	5.733	1.33	5.05
1 (linear = OBPI)	94.72%	103.71	9.934	1.4345	5.59
2 (convex)	83.72%	104.88	22.06	1.5971	6.17
3 (convex)	72.33%	106.31	36.52	1.98	8.82
4 (convex)	60.86%	108.06	55.07	2.61	14.02
Asset S	0%	104.23	14.93	0.35	3.13

The call-power moments are increasing functions of m, whereas the guaranteed level is decreasing w.r.t. the parameter m.

The following two figures correspond respectively to the pdf and cdf of call-power options.

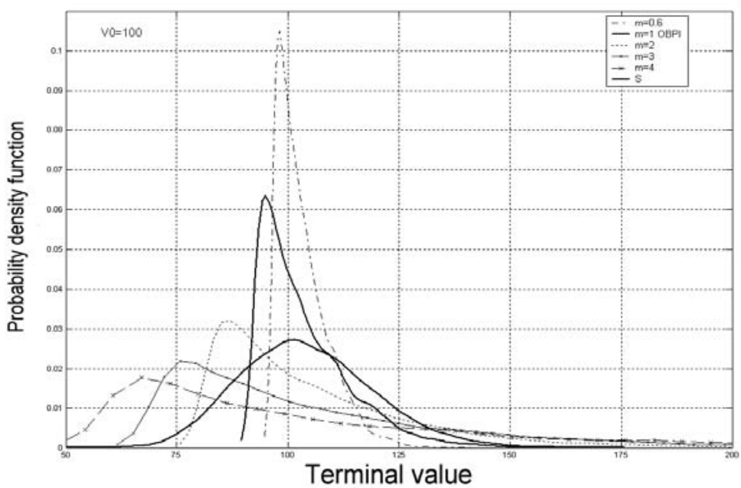


FIGURE 9.4: Call-power option pdf

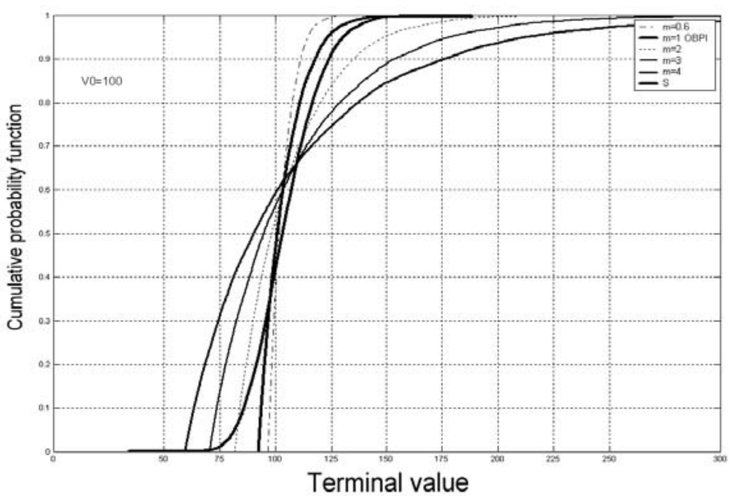


FIGURE 9.5: Call-power option cdf

Note that all cdf graphs intersect each other, which means that no stochastic dominance exists. Besides, the higher the possible gain, the smaller the guaranteed level.

Note that, for the case $h_c(S_T) = aS_T + b$, the portfolio value V_T also has the following form:

$$V_T = aS_T + b + \left(\underbrace{d.S_T^m - aS_T}_{Q(S)} - b\right)^+.$$

The study of such portfolio payoffs (see for example [417]) requires the examination of polynomial option properties.

As shown by Macovschi and Quittard-Pinon [370], a polynomial option can be decomposed as a sum of power options. Indeed, we have to search for roots of the characteristic polynomial function associated to the payoff function. Let $Q(S) = \sum_{j=1}^n a_j S^j$ be a polynomial function with degree n, and let b be a positive scalar such that the polynomial function $P(S) = Q(S) - b$ has exactly p non-negative roots $\lambda_1, \lambda_2, ..., \lambda_p$ with $\lambda_1 < \lambda_2 <, ..., < \lambda_p$. Then the European call with strike b can be valued from the relation:

$$\max\left(Q(S) - b, 0\right) \Longrightarrow \quad C_Q(b) = \sum_{j=1}^n a_j \left(\sum_{k=1}^p (-1)^{k+1} C_j\left((\lambda_k)^j\right)\right), \quad (9.11)$$

where $C_j(H)$ is the power-j option, with strike H and maturity T.

The following example illustrates the possible payoffs of such portfolios according to the values of parameters d and m.

Example 9.2

Assume that the financial parameters are given by:

$$\mu = 0.1,\ \sigma = 0.2 S_0 = 100 \text{ and } r = 3\%.$$

Assume also that the investor's characteristics are the following ones:

$$V_0 = 100, T = 5, \text{ and } a = 0.7,\ b = 90,\ d = 8.7253.10^{-11},\ m = 5.8333.$$

The portfolio value with guarantee constraint is given by:

$$V_T^{**} = 0.7S_T + 90 + \max\left(8.7253.10^{-11}.S_T^{5.8333} - 0.7S_T - 90; 0\right).$$

The characteristic polynomial function associated to the payoff is equal to:

$$P(S) = 8.7253.10^{-11}.S_T^{5.8333} - 0.7S_T - 90.$$

It has only one positive root, $\lambda = 129.1958$.

From Equation (9.11), we deduce the polynomial option value:

$$\begin{aligned} C_Q(90) &= \sum_{j=1}^{n} a_j C_j \left((\lambda) ,^j \right) \\ &= -0.7.C_1(129.1958) + 8.7253.10^{-11}.C_3 \left(129.1958^{5.8333}\right), \end{aligned}$$

where $C_1(K)$ and $C_3(K)$ are respectively equal to the standard Black-Scholes call value and the Black-Scholes value of the power-5.8333 option with strike K.

Recall that the Black-Scholes power-m option value is given by:

$$\begin{aligned} V_0 &= e^{-r(1-m)T} \left[\widetilde{S}_0 N(\widetilde{d}_1) - K e^{-\widetilde{r}T} N(\widetilde{d}_2) \right], \\ &= S_0^m e^{\left[\left(r+\frac{1}{2}\sigma^2 m\right)T(m-1)\right]} N(\widetilde{d}_1) - K e^{-rT} N(\widetilde{d}_2), \end{aligned}$$

with

$$\widetilde{d}_1 = \frac{\ln\left(\frac{\alpha S_0^m e^{\frac{1}{2}\sigma^2 m(m-1)T}}{K}\right) + \left(mr + \frac{1}{2}(m\sigma)^2\right)T}{m\sigma\sqrt{T}},$$

and

$$\widetilde{d}_2 = \widetilde{d}_1 - m\sigma\sqrt{T}.$$

Then, we get:

$$C_Q(90) = 18.9.$$

Thus, we deduce the initial portfolio value which allows the investors to receive such guarantee at maturity:

$$V_0^{**} = 0.7 S_0 + 90.e^{-rT} + C_Q(90) = 166.37.$$

This investor is sure to recover 54.09% of his initial investment, at maturity. □

The following figure indicates the terminal values of the portfolio as function of the risky asset values S_T.

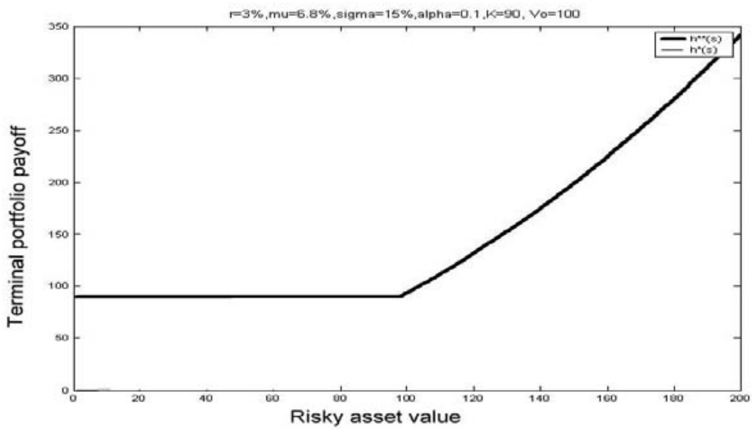


FIGURE 9.6: Convex case with linear constraints

An example of concave profile with linear constraint:

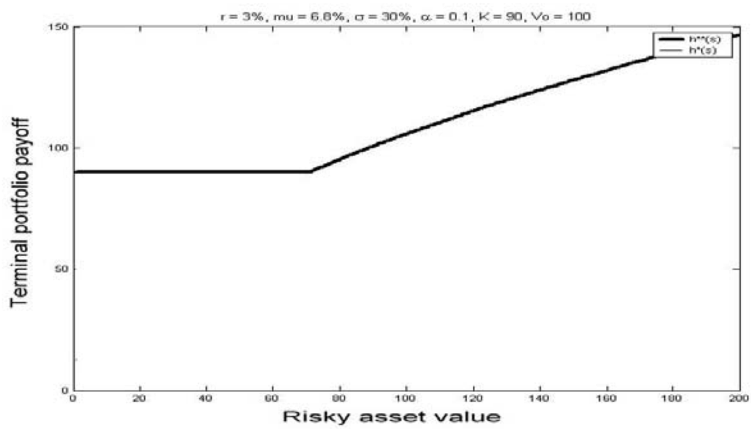


FIGURE 9.7: Concave case with linear constraints

9.1.2.2 Other possible extensions

Obviously, other options can be introduced to provide a percentage of the potential market rise. Indeed, any portfolio with terminal value V_T such that:

$$V_T = K + H_T, \tag{9.12}$$

where H_T is a positive random variable is always above the level K.

The choice of a particular payoff H_T may depend on:

- Market predictions: rise or drop of the financial market, volatility levels, *etc.*
- The type of risky assets: financial index, hedge fund, *etc.*
- The insurance cost associated to each chosen derivative: lookback options, corridor options, *etc.*

REMARK 9.3 As seen in the next chapter, from a theoretical point of view, it may be optimal to use options on a power of the risky asset.

$\square$

From Relations (9.1 and 9.2), the reduction of the insurance cost can allow the investor to get a higher percentage of the market rise.

Among possible options, consider for example the capped Call C_T^{capp}:

$$C_T^{capp} = K + \mathrm{Min}(\mathrm{Max}(S_T - K, 0); K'), \tag{9.13}$$

which always provide a guaranteed amount equal to K, is equal to the risky asset value S_T when $K \leq S_T \leq K'$, but remains constant equal to $K + K'$ for high values of S_T.

This allows reduction of the insurance cost and is profitable if the value S_T remains below K'.

However, options must often be synthesized by a dynamic hedging strategy when they are not available on the financial market.

Therefore, standard problems appear, such as imperfect hedging, transaction costs, and rebalancing strategy, which generally increase the insurance cost.

9.2 The Constant Proportion Portfolio Insurance

The CPPI, introduced by Perold [406], uses a simplified strategy to allocate assets dynamically over time. The investor starts by setting a floor equal to the lowest acceptable value of the portfolio. Then, he computes the cushion as the excess of the portfolio value over the floor and determines the amount allocated to the risky asset by multiplying the cushion by a predetermined multiple. Both the floor and the multiple are functions of the investor's risk tolerance and are exogenous to the model. The total amount allocated to the risky asset is known as the exposure. The remaining funds are invested in the reserve asset, usually T-bills.

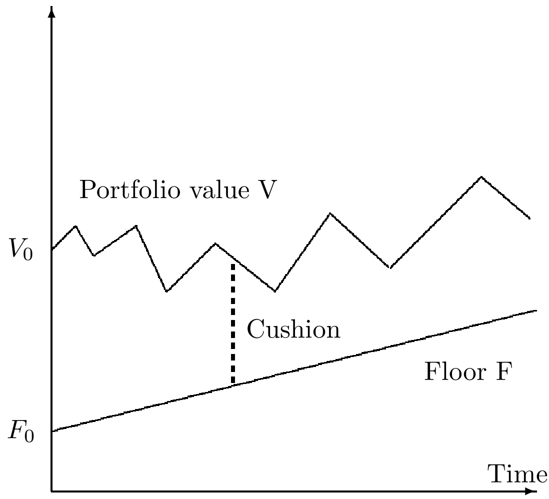


FIGURE 9.8: Portfolio value and cushion

The higher the multiple, the more the investor will participate in a sustained increase in stock prices. Nevertheless, the higher the multiple, the faster the portfolio will approach the floor when there is a sustained decrease in stock prices. As the cushion approaches zero, exposure approaches zero too. In continuous time, this keeps the portfolio value from falling below the floor. Portfolio value will fall below the floor only when there is a very sharp drop in the market before the investor has a chance to trade.

In what follows, we refer mainly to Prigent [413].

9.2.1 The standard CPPI method

The CPPI method consists of managing a dynamic portfolio so that its value is above a floor F at any time t. The value of the floor gives the dynamical insured amount. It is assumed to evolve as a riskless asset, according to:

$$dF_t = F_t r dt. \tag{9.14}$$

Obviously, the initial floor F_0 is less than the initial portfolio value V_0^{CPPI}. The difference $V_0^{CPPI} - F_0$ is called the cushion, denoted by C_0. Its value C_t at any time t in $[0, T]$ is given by :

$$C_t = V_t^{CPPI} - F_t. \tag{9.15}$$

Denote by e_t the exposure, which is the total amount invested in the risky asset. The standard CPPI method consists of letting

$$e_t = mC_t, \tag{9.16}$$

where m is a constant called the multiple. Note that the interesting case for portfolio insurance corresponds to $m > 1$, that is, when the payoff function is convex and can provide significant percentage of the market rise.

Assume that the risky asset price process $(S_t)_t$ is a diffusion with jumps:

$$dS_t = S_{t-}[\mu(t, S_t)dt + \sigma(t, S_t)dW_t + \delta(t, S_t)d\gamma], \tag{9.17}$$

where $(W_t)_t$ is a standard Brownian motion, independent from the Poisson process with measure of jumps γ.

In particular, this means that:

- The sequence of random times $(T_n)_n$ corresponding to jumps satisfies the following properties: the interarrival times $T_{n+1} - T_n$ are independent and have the same exponential distribution with parameter denoted by λ.
- The relative jumps of the risky asset $\frac{\Delta S_{T_n}}{S_{T_n}}$ are equal to $\delta(T_n, S_{T_n})$. They are supposed to be strictly higher than -1 (in order for the price S to be stricly positive).

We deduce the portfolio value:

PROPOSITION 9.1
i) The value of this portfolio V_t^{CPPI} at any time t in the period $[0,T]$ is given by:

$$V_t^{CPPI}(m, S_t) = F_0.e^{rt} + C_t, \tag{9.18}$$

where the cushion value C_t is equal to:

$$C_t = C_0 exp\left((1-m)rt + m\left[\int_0^t (\mu - 1/2m\sigma^2)(s, S_s)ds + \int_0^t \sigma(s, S_s)dW_s\right]\right)$$
$$\times \prod_{0 \leq T_n \leq t} (1 + m\ \delta(T_n, S_{T_n})). \tag{9.19}$$

ii) When the risky asset price has no jump ($\delta = 0$), and the coefficients $\mu(.,.)$ and $\sigma(.,.)$ are constant, we deduce the following standard formula:

$$V_t^{CPPI}(m, S_t) = F_0.e^{rt} + \alpha_t.S_t^m, \tag{9.20}$$

where

$$\alpha_t = \left(\frac{C_0}{S_0^m}\right) \exp[\beta t] \text{ and } \beta = \left(r - m\left(r - \frac{1}{2}\sigma^2\right) - m^2\frac{\sigma^2}{2}\right). \tag{9.21}$$

PROOF - The portfolio value of the CPPI strategy is given by:

$$dV_t^{CPPI} = (V_t^{CPPI} - e_t)\frac{dB_t}{B_t} + e_t\frac{dS_t}{S_t}.$$

The CPPI strategy is based on the following relations: $V_t^{CPPI} = C_t + F_t$, $e_t = mC_t$, and the floor value satisfies $dF_t = rdt$. Therefore, the cushion value is given by:

$$\begin{aligned}
dC_t &= d(V_t^{CPPI} - F_t), \\
&= (V_t^{CPPI} - e_t)\tfrac{dB_t}{B_t} + (e_t)\tfrac{dS_t}{S_t} - dF_t, \\
&= (C_t + F_t - mC_t)\tfrac{dB_t}{B_t} + (mC_t)\tfrac{dS_t}{S_t} - dF_t, \\
&= (C_t - mC_t)\tfrac{dB_t}{B_t} + (mC_t)\tfrac{dS_t}{S_t}, \\
&= C_t[r + m\,(\mu(t, S_t) - r)\,dt + m\sigma(t, S_t)dW_t + m\delta(t, S_t)d\gamma].
\end{aligned}$$

Consequently, the cushion value is a stochastic exponential (the so-called Doléans-Dade exponential as shown in Appendix B).

Thus, the cushion value C_t at any time t is given by:

$$C_0 exp\left((1-m)rt + m\left[\int_0^t (\mu - 1/2m\sigma^2)(s, S_s)ds + \int_0^t \sigma(s, S_s)dW_s\right]\right)$$
$$\times \prod_{0 \leq T_n \leq t} (1 + m\ \delta(T_n, S_{T_n})).$$

When the risky asset price has no jump and the coefficients $\mu(.,.)$ and $\sigma(.,.)$ are constant, we have:

$$C_t = C_0 exp[((m(\mu - r) + r - \frac{m^2\sigma^2}{2})t + m\sigma W_t)]. \tag{9.22}$$

Then, using the relation $S_t = S_0 \exp\left[\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right)t\right]$, we deduce:

$$W_t = \frac{1}{\sigma}\left[\ln\left(\frac{S_t}{S_0}\right) - \left(\mu - \frac{1}{2}\sigma^2\right)t\right].$$

Therefore, substituting W_t in the expression of C_t, we have:

$$C_t\left(m, S_t\right) = C_0\left(\frac{St}{S_0}\right)^m \exp\left[\left(r - m\left(r - \frac{1}{2}\sigma^2\right) - m^2\frac{\sigma^2}{2}\right)t\right] = \alpha_t.S_t^m,$$

where

$$\alpha_t = \left(\frac{C_0}{S_0^m}\right)\exp\left[\beta t\right] \text{ and } \beta = \left(r - m\left(r - \frac{1}{2}\sigma^2\right) - m^2\frac{\sigma^2}{2}\right).$$

Finally, the CPPI portfolio value is given by:

$$V_t^{CPPI}\left(m, S_t\right) = F_0.e^{rt} + \alpha_t.S_t^m.$$

□

REMARK 9.4 Consequently, the guarantee constraint is satisfied as soon as the relative jumps are such that:

$$\delta(T_n, S_{T_n}) \geq -1/m. \tag{9.23}$$

Thus, when the risky asset jumps are higher than a constant, then the condition $0 \leq m \leq -1/d$ allows the positivity of the cushion value. For example, if d is equal to -20%,we have $m \leq 5$. If d is equal to -10%,we have $m \leq 10$. Note that these upper bounds on the multiple do not depend on the probability distribution of the jump times ΔS_{T_n}. □

REMARK 9.5 - When the risky asset price is a geometric Brownian motion ($\delta = 0$ and μ and σ are constant), the cushion value is given by:

$$C_t = C_0 e^{m\sigma W_t + [r + m(\mu - r) - \frac{m^2\sigma^2}{2}]t} \ with \ C_0 = V_0 - P_0. \tag{9.24}$$

In that case, the portfolio and cushion values are path independent. They have lognormal distributions, up to the deterministic floor value F_T for the portfolio. The volatility is equal to $m\sigma$, and the instantaneous mean is given by $r + m(\mu - r)$. This illustrates the leverage effect of the multiple m: the higher the multiple, the higher the excess return but also the volatility. Since we have $V_t^{CPPI}\left(m, S_t\right) = F_0.e^{rt} + \alpha_t.S_t^m$, the portfolio profile is convex as soon as the multiple m is higher than 1.

□

9.2.1.1 Lévy process case

Assume that $\delta(.)$ is not equal to 0, and μ and σ are constant. Suppose also that the jumps are iid with probability distribution $H(dx)$, with finite mean $E[\delta(T_1, S_{T_1})]$ denoted by b and $E[\delta^2(T_1, S_{T_1})] < \infty$ equal to c. The logarithmic return of the risky asset S is a Lévy process (see Appendix B). Then, we deduce:

PROPOSITION 9.2
- The first two moments of the portfolio values are given by:

$$\begin{cases} \mathbb{E}[V_t] & = (V_0 - P_0)e^{[r+m(\mu+b\lambda-r)]t} + P_0 e^{rt}, \\ Var[V_t] & = (V_0 - P_0)^2 e^{2[r+m(\mu+b\lambda-r)]t}[e^{m^2(\sigma^2+c\lambda)t} - 1]. \end{cases} \tag{9.25}$$

- The pdf of the cushion C_t is determined as follows. Let $g_{0,t}$ denote the pdf without jump. The function $g_{0,t}$ is defined by:

$$g_{0,t}(x) = \frac{I_{x>0}}{x\sqrt{2\pi\sigma^2 m^2 t}} e^{-\frac{1}{2\sigma^2 m^2 t}\left(ln\left[\frac{x}{C_0}\right] - t(r+m(\mu-r)-\frac{m^2\sigma^2}{2})\right)^2}. \tag{9.26}$$

When jumps can occur, we have:

$$g_{Ct}(x) = \sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!}\int_{I\!R^n} g_{0,t}(\frac{x}{\prod_{i\leq n}(1+my_i)})\frac{K^n(dy_1,...,dy_n)}{\prod_{i\leq n}(1+my_i)}, \tag{9.27}$$

where H^n denotes the joint distributon of the relative jumps $\delta(T_i, S_{T_i}), i \leq n$. Since they are assumed to be independent with the same distribution H, we deduce the relation:

$$H^n(dy_1, ..., dy_n) = \otimes H(dy_i). \tag{9.28}$$

Consequently, the pdf of the portfolio value is given by:

$$f_{Vt}(x) = g_{Ct}(x - P_0 e^{rt}). \tag{9.29}$$

REMARK 9.6 From the mean-variance point of view, note that both mean and variance of the CPPI portfolio value are increasing functions of the multiple m and decreasing w.r.t. the initial floor value F_0. It is not possible to optimize w.r.t. the multiple m according to Markowitz criterion. Indeed, for a fixed mean level L the intial floor value F_0 is a function of the multiple given by:

$$F_0(m) = \frac{LV_0 - V_0 e^{[r+m(\mu+b\lambda-r)]t}}{e^{rt}[1 - e^{m(\mu+b\lambda-r)t}]}, \tag{9.30}$$

which implies

$$Var[V_t/V_0] = (e^{rt} - L)^2 \frac{[e^{m^2(\sigma^2+\lambda c)t} - 1]}{[1 - e^{-m(\mu+b\lambda-r)t}]^2}, \tag{9.31}$$

which is a function of the multiple not having a real minimum.

In fact, the expectation of gains is not the first goal: the guarantee level is crucial. When it has been determined, the multiple allows to adjust the percentage of anticipated profit when the financial market rises. □

The mean of the portfolio return is an increasing function of the multiple m. However, as mentioned previously in relation (9.23), the risky asset discontinuities lead to upper bounds on the multiple m.

In order to reduce this constraint, another weaker guarantee condition can be considered which can be based on Value-at-Risk or Expected Shortfall criterion. For example, consider the VaR type condition:

$$\mathbb{P}[C_t \geq 0, \forall t \leq T] \geq 1 - \epsilon, \tag{9.32}$$

where ϵ is "small." This condition is equivalent to:

$$\mathbb{P}[\forall t \leq T, \frac{\Delta S_t}{S_t} \geq \frac{-1}{m}] \geq 1 - \epsilon. \tag{9.33}$$

Denote by H the cdf of relative jumps and asume that it is strictly increasing. Then, we deduce:

PROPOSITION 9.3

The condition

$$\mathbb{P}[C_t \geq 0, \forall t \leq T] \geq 1 - \epsilon$$

is equivalent to the following condition on the multiple m :

$$m \leq \frac{-1}{H^{(-1)}\left(\frac{1}{\lambda T} ln(\frac{1}{1-\epsilon})\right)}. \tag{9.34}$$

PROOF The VaR guarantee condition fo a given threshold $1 - \epsilon$ is the following:

$$\mathbb{P}[C_t \geq 0, \forall t \leq T] \geq 1 - \epsilon. \tag{9.35}$$

The cushion value provided in relation (9.19) shows that the variation due to jumps is equal to $\prod_{0 \leq T_n \leq t}(1 + m\ \delta(T_n, S_{T_n})$. This term must remain positive with a probability equal to $1 - \epsilon$ at any time t. Therefore, each term in this product must be positive. This is equivalent to:

$$\mathbb{P}\left[\cap_{T_n \leq T}\left\{\delta(T_n, S_{T_n}) \geq \frac{-1}{m}\right\}\right] \geq 1 - \epsilon. \tag{9.36}$$

Denote by N_T the number of jumps before maturity T. Since in the Lévy case jumps are independent from their occurence times, we deduce:

$$\mathbb{P}\left[\cap_{T_n \leq T}\left\{\delta(T_n, S_{T_n}) \geq \frac{-1}{m}\right\}\right] =$$
$$\sum_k \mathbb{P}\left[\cap_{n\leq k}\left\{\delta(T_n, S_{T_n}) \geq \frac{-1}{m}\right\}\right]\mathbb{P}\left[N_T = k\right]. \tag{9.37}$$

The random variable N_T, which counts the number of jumps during the time period $[0, T]$, is a Poisson distribution with parameter λT. Therefore, the VaR guarantee constraint at the threshold $1 - \epsilon$ is equivalent to:

$$m \leq \frac{-1}{H^{(-1)}\left(\frac{1}{\lambda T} ln(\frac{1}{1-\epsilon})\right)}. \tag{9.38}$$

□

REMARK 9.7 The latter condition (9.38) provides an upper bound on the multiple which is obviously higher than the opposite of the infimum d of the range of the distribution H. Note also that this upper bound is a decreasing function of the jump intensity λ. For high values of λ, the upper bound converges to $-1/d$. □

9.2.1.2 Discrete-time case

Suppose now that the investor trades according a discrete-time grid: t_k, $k \leq n$. For example, he wants to limit transaction costs or portfolio assets are not sufficiently liquid. In this framework, the CPPI portfolio value can be determined as follows

Denote by X_k the opposite of the arithmetical return of the risky asset between t_{k-1} and t_k. We have:

$$X_k = -\frac{S_{t_k} - S_{t_{k-1}}}{S_{t_{k-1}}}. \tag{9.39}$$

Consider the maximum M of these values. For each l, define:

$$M_l = Max(X_1, ..., X_l).$$

Denote by V_k the portfolio value at time t_k. The guarantee constraint is to keep the portfolio value V_k above the floor P_k. The exposure e_k invested in the risky asset S_k is equal to mC_k, where the cushion value C_k is equal to $V_k - P_k$. The remaining amount $(V_k - e_k)$ is invested in the riskless asset with return r_k on the time period $[t_k, t_{k+1}]$.

Therefore, the dynamic evolution of the portfolio value is given by (similar to the continuous-time case):

$$V_{k+1} = V_k - e_k X_{k+1} + (V_k - e_k) r_{k+1}, \tag{9.40}$$

where the cushion value is defined by:

$$C_{k+1} = C_k \left[1 - m\ X_{k+1} + (1-m) r_{k+1}\right]. \tag{9.41}$$

Since at any time t_k, the cushion value must be positive, we get, for any $k \leq n$,

$$-mX_k + (1-m) r_k \geq -1.$$

Besides, since r_k is relatively small, the previous inequality is "approximately" equivalent to:

$$\forall k \leq n, X_k \leq \frac{1}{m} \text{ then also } M_n = Max(X_k)_{k \leq n} \leq \frac{1}{m}.$$

We deduce that the guarantee condition is satisfied as soon as the multiple m is smaller than d (equal to the infimum f the range of the distribution of the random variables X_k).

For a VaR guarantee constraint at the level $(1-\epsilon)$ on the time period $[0, T]$:

$$\mathbb{P}[C_t > 0, \forall t \in [0, T]] \geq 1 - \epsilon,$$

the maximum M_n of the values X_k at times t_k during $[0, T]$ must satisfy:

$$\mathbb{P}[\forall t_k \in [0, T], X_k < \frac{1}{m}] = \mathbb{P}[M_n < \frac{1}{m}] \geq 1 - \epsilon. \tag{9.42}$$

According to the distributions of the random variables X_k, upper bounds on the multiple are deduced.

Suppose for example that the random variables X_k are iid with cdf H which is assumed to have an inverse function $H^{(-1)}$.

Then, the following condition must be satisfied by the multiple m:

$$m < \frac{1}{F^{-1}(1 - \sqrt[n]{1-\epsilon})}. \tag{9.43}$$

9.2.1.3 Random rebalancing case

When the cushion rises due to market fluctuations, the exposure may be close to the maximum amount that the investor wants to invest in the risky asset. When the exposure is below this limit, the investor can have a *tolerance* w.r.t. these market variations. It means that he can fix a percentage of market drops and rises such that, when market fluctuations are above this level, he rebalances his portfolio.

Consider for example a lower bound $\underline{m}$ and an upper bound $\bar{m}$ on the multiple m^*. The investor begins by choosing an initial floor F_0, a quantity θ_0^S, invested in the risky asset S and a quantity θ_0^B invested in the riskless asset B.
From initial conditions, we deduce:

$$\theta_0^S = \frac{m^*(V_0 - P_0)}{S_0}. \tag{9.44}$$

The portfolio is rebalanced as soon as the ratio $\frac{e_t}{C_t}$ is smaller than $\underline{m}$ or higher than $\bar{m}$. If $\theta_0^B < \frac{P_0}{B_0}$, then, for the geometric Brownian case, the condition

$$\underline{m} \leq \frac{e_t}{C_t} \leq \bar{m} \tag{9.45}$$

is equivalent to:

$$A \leq X_t \leq B, \tag{9.46}$$

where $(X_t)_t$ is a Brownian motion with drift, defined by:

$$X_t = (\mu - r - 1/2\sigma^2)t + \sigma W_t,$$

and the paramaters A and B are given by:

$$\begin{cases} A = Ln\left(\frac{\bar{m}}{\bar{m}-1}\frac{(P_0-\theta_0^B B_0)}{m^*(V_0-P_0)}\right) = Ln\left(\frac{\bar{m}}{\bar{m}-1}\frac{(m^*-1)}{m^*}\right), \\ B = Ln\left(\frac{\underline{m}}{\underline{m}-1}\frac{(P_0-\theta_0^B B_0)}{m^*(V_0-P_0)}\right) = Ln\left(\frac{\underline{m}}{\underline{m}-1}\frac{(m^*-1)}{m^*}\right). \end{cases}$$

The conditional distribution of the rebalancing times is characterized by the exist time of the process $(X_t)_t$ from the corridor $\{A, B\}$. This probability distribution is deduced from the trivariate distribution of the maximum, minimum, and terminal value of the Brownian motion (see Revuz and Yor [423] or Borodine and Salminen [85]).

The pdf of this joint distribution with a constant drift ρ is defined by, for any x in $\{A, B\}$:

$$g(x, A, B) = exp[\frac{\rho x}{\sigma^2} - \frac{\rho^2 t}{2\sigma^2}]\times$$

$$\sum_{n=-\infty}^{+\infty} \frac{1}{\sigma\sqrt{t}}\left(\phi(\frac{x - 2n(B-A)}{\sigma\sqrt{t}}) - \phi(\frac{x - 2n(B-A) - 2A}{\sigma\sqrt{t}})\right), \tag{9.47}$$

where ϕ is the pdf of the standard Gaussian distribution and N its cdf.

If $A < 0$ and $B > 0$, then the distribution of the first passage time T_1 is given by:

$$\mathbb{P}[T_1 \leq t] = 1 - \mathbb{P}[Max_{s\leq t} X_s \leq B, Min_{s\leq t} X_s \geq A] = 1-$$
$$\sum_{n=-\infty}^{+\infty} e^{2n\rho(B-A)/\sigma^2}[N(\tfrac{B-\rho t-2n(B-A)}{\sigma\sqrt{t}}) - N(\tfrac{A-\rho t-2n(B-A)}{\sigma\sqrt{t}})]$$
$$-e^{2A\rho/\sigma^2}[N(\tfrac{B-\rho t-2n(B-A)-2A}{\sigma\sqrt{t}}) - N(\tfrac{A-\rho t-2n(B-A)-2A}{\sigma\sqrt{t}})].$$

After the firsttime T_1, the new portfolio value is determined from the following relations: The new initial floor is equal to $P_0 e^{rT_1}$. The quantities $\theta_{T_1}^S$ and $\theta_{T_1}^B$ respectively invested in S and B are determined from rebalancing conditions. We have in particular:

$$\theta_{T_1}^S = \frac{m^*(V_{T_1} - P_0 e^{rT_1})}{S_{T_1}}. \tag{9.48}$$

From the Lévy property, the distribution of the new interarrival time $T_2 - T_1$ is deduced from the previous one, by stopping all processes at time T_1. Note that when jumps can occur but the logarithmic return is still a Lévy process, the previous cdf can be defined from infinite expansions.

9.2.2 CPPI extensions

The CPPI method is based on an exposure e which is a simple linear function w.r.t. the cushion. It can be extended by introducing a more general exposure function $e(t,x)$, defined on $[0,T] \times \mathbb{R}^+$, which is assumed to be positive and continuous and has the following form:

$$e_t = e(t, C_t). \tag{9.49}$$

Consequently, the cushion is the solution of the following SDE:

$$dC_t = \alpha(t,S_t,C_t)dt + \beta(t,S_t,C_t)dW_t + \gamma(t,S_t,C_t)d\mu,$$
$$\text{with}$$
$$\begin{cases} \alpha(t,S_t,C_t) = rC_t + e(t,C_t))[a(t,S_t) - r], \\ \beta(t,S_t,C_t) = e(t,C_t)\sigma(t,S_t), \\ \gamma(t,S_t,C_t) = e(t,C_t)\delta(t,S_t). \end{cases} \tag{9.50}$$

The positivity of the cushion is controlled by an appropriate choice of the function $e(.,.)$:

- If the cushion is null then the exposure $e(t,0)$ must be equal to 0.
- If the relative jumps $\frac{\Delta S}{S-}$ are higher than a fixed constant d (negative), then for any (t,x), we must have $e(t,x) \leq -\frac{1}{d}x$.

REMARK 9.8 The implications of this method can be analyzed by considering conditions on buying and selling, probability distribution of the cushion value, *etc.* Indeed, a more general exposure function allows for a better performance since the portfolio manager has more available parameters to adjust the portfolio profile.

Such a choice is compatible with tactical allocation and can be applied with fixed income instruments or hedge funds. For example, the multiple m can be dynamically chosen to take account of the implicit volatility of options of the financial market or other factors.

We can also impose a more general guarantee condition

$$V_T \geq F_T$$

at maturity T, where F_T is a contingent claim such that for example,

$$F_T > P_0 e^{rT}.$$

Sometimes, the investor wants to keep until maturity some part of the portfolio gains. For this purpose, the *Time Invariant Protection Insurance* (TIPP), introduced by Ested and Kritzman [213] can be used. This method allows to have at any time t:

$$V_t \geq kMax(F_t,\ sup_{s \leq t} V_s),\ \text{with } 0 < k < 1. \tag{9.51}$$

In particular, it means that the investor does not want to lose more than a given percentage of the maximum of previous portfolios values. Denote:

$$X_t = Max(F_t, sup_{s \leq t} V_s).$$

In that case, the extended CPPI method is based on the new floor kX_t and the new exposure:

$$e_t = m\left(V_t - kX_t\right).$$

We can also consider a more general function of $Max(F_t, sup_{s \leq t} V_s)$:

$$e_t = e\left(t, Max(F_t, sup_{s \leq t} V_s)\right).$$

As presented in the next chapter, we can weaken the constraint $V_T \geq F_T$ by imposing only this condition at a given probability threshold. The insurance cost is reduced but the guarantee is no longer sure.

□

9.3 Comparison between OBPI and CPPI

In what follows we refer to Bertrand and Prigent [60]. We assume that the same initial amount V_0 is invested at time 0, and also that the same guarantee K holds at maturity. We suppose also that the risky asset price follows a geometric Brownian motion.

9.3.1 Comparison at maturity

9.3.1.1 Comparison of the payoff functions

Is it possible that the payoff function of one of these two strategies lies above the other for all S_T values? Since the initial investments are equal ($V_0^{OBPI} = V_0^{CPPI}$), the absence of arbitrage implies the following result.

PROPOSITION 9.4
Neither of the two payoffs is greater than the other for all terminal values of the risky asset. Therefore, the two payoff functions intersect one another.

Figure 9.9 illustrates a numerical example with typical values for the financial market parameters: $\mu = 10\%$, $\sigma = 20\%$, $T = 1$, $r = 5\%$, $K = S_0 = 100$. Note that as m increases, the payoff function of the CPPI becomes more convex.

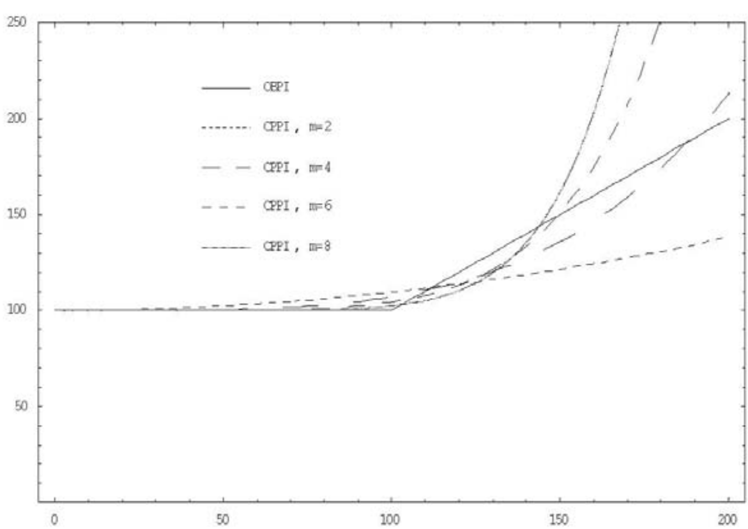


FIGURE 9.9: CPPI and OBPI payoffs as functions of S

For this example, the two curves intersect one another for the different values of m considered ($m = 2$, $m = 4$, $m = 6$, and $m = 8$).

CPPI performs better for large fluctuations of the market, while OBPI performs better in moderate bullish markets.

9.3.1.2 Comparison with the stochastic dominance criterion

The first-order stochastic dominance allows us to take account of the risky dimension of the terminal payoff functions for both methods.

Recall that a random variable X stochastically dominates a random variable Y at the first order ($X \succ Y$) if and only if the cumulative distribution function of X, denoted by F_X, is always below the cumulative distribution function F_Y of Y.

PROPOSITION 9.5
Neither of the two strategies stochastically dominates the other at first order.

9.3.1.3 Comparison of the expectation, variance, skewness, and kurtosis

Since option payoffs are not linear w.r.t. the underlying risky asset, the mean-variance criterion is not always justified. Thus, we examine simultaneously the first four moments and the semi-variance of the rates of portfolio returns R_T^{OBPI} and R_T^{CPPI}.

PROPOSITION 9.6
The equality of return expectations of both strategies,

$$\mathbb{E}[R_T^{OBPI}] = \mathbb{E}[R_T^{CPPI}],$$

leads to a unique value for the multiple, denoted by $m^*(K)$, for any fixed guaranteed amount K. In the Black and Scholes framework, this multiple is equal to:

$$m^*(K) = 1 + \left(\frac{1}{(\mu - r)T}\right) \ln\left(\frac{C(0, S_0, K, \mu)}{C(0, S_0, K, r)}\right), \tag{9.52}$$

where $C(0, S_0, K, x)$ is the Black-Scholes value of the call option, and x denotes all possible values of the riskless rate.

PROPOSITION 9.7
For any parametrization of the financial markets (S_0, K, μ, σ, r), there exists at least one value for m such that the OBPI strategy dominates (is dominated), in a mean-variance (mean-semivariance) sense, (by) the CPPI one.

The following example gives an illustration with the previous values of the parameters.

The multiple m, solution of $\mathbb{E}[R_T^{OBPI}] = \mathbb{E}[R_T^{CPPI}]$, is equal to 5.77647.

The next table contains the first four moments and the semi-volatility for the OBPI with an at-the-money call, and for the corresponding CPPI with this particular value of the multiple.

TABLE 9.2: Comparison of the first four moments and semi-volatility

	OBPI	CPPI
expectation	8.61176 %	8.61176 %
volatility	16.8625 %	23.2395 %
semi-volatility	9.1676%	7.7666%
relative skewness	1.49114	9.70126
relative kurtosis	5.4576	357.73

- The OBPI dominates the CPPI in a mean-variance sense, but is dominated by the CPPI if semi-volatility is considered. This is confirmed by the relative skewness.
- Nevertheless, the CPPI has a higher positive relative skewness than the OBPI, and should be preferred to OBPI for this criterion.
- However, CPPI relative kurtosis is much higher than the OBPI one. This is due to the dominance of the CPPI payoff for small and high values of the risky asset S, as shown in Figure 9.9.
- Note that, here, owing to the insurance feature, kurtosis arises mainly in the right tail of the distribution.

9.3.1.4 Comparison of "quantiles"

Since the distributions to be compared are strongly asymmetrical, the study of the moments is not sufficient. The whole distribution has to be considered. The next figure loosely illustrates the situation, where both payoff functions and risky asset density are depicted.

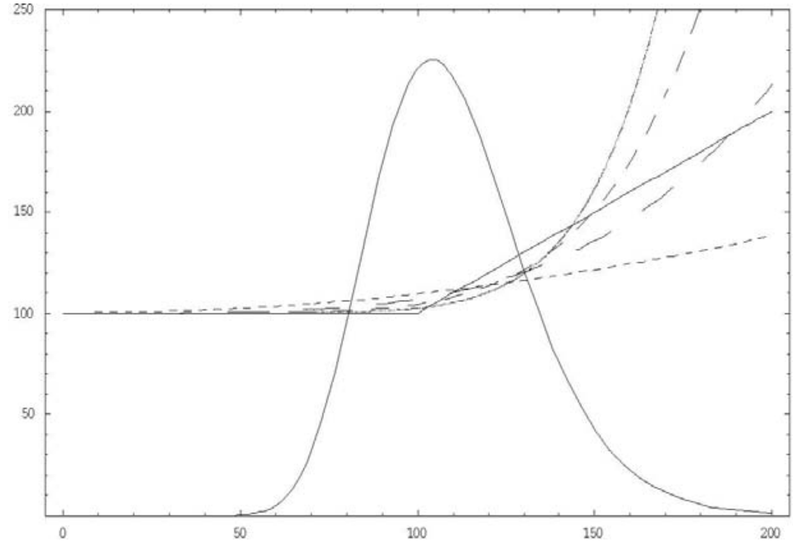


FIGURE 9.10: CPPI and OBPI payoffs and probability of S

To examine the effect of probabilities, we study the distribution of the quotient of the CPPI value to the OBPI one. The plot of the cumulative distribution function of $\frac{V_T^{OBPI}}{V_T^{CPPI}}$ for different values of K is:

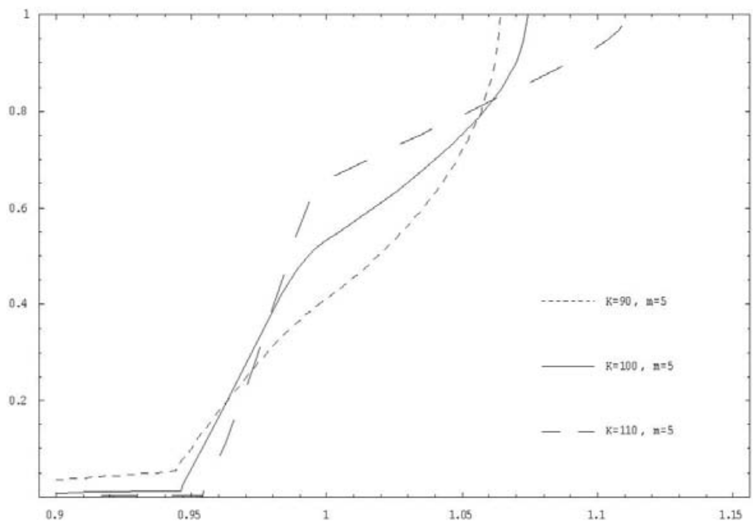


FIGURE 9.11: Cumulative distribution of $\frac{V_T^{OBPI}}{V_T^{CPPI}}$

This figure shows in particular that:

- For the at-the-money call ($K = 100$ and thus $p = 94.72\%$), the probability that the CPPI portfolio value is higher than the OBPI one is approximately 0.5, meaning that neither of the strategies "dominates" the other.
- This is no longer true for $K = 90$ (thus $p = 87.97\%$), where the probability that the CPPI portfolio value is above the OBPI one is about 0.4.
- For $K = 110$ (thus $p = 99.39\%$), this probability takes the value 0.7.

REMARK 9.9

- This arises because the probability of exercising the call decreases with the strike. Recall that the strike K is an increasing function of the insured percentage p of the initial investment. Thus, as the guaranteed percentage p rises, the CPPI method is more desirable than the OBPI method.
- Notice that, for in- and out-of-the-money calls, extreme values of the quotient are more likely to appear:

  - On the one hand, the CPPI portfolio value can be at least equal to 106% of the OBPI portfolio value with probability 5% (respectively about 0%) when $K = 90$ (respectively $K = 110$).

  - On the other hand, the CPPI portfolio value can be at most equal to 94% of the OBPI portfolio value with probability about 0% (respectively 18%) when $K = 90$ (respectively $K = 110$).
- The same qualitative results are obtained for other usual values of the multiple (m between 2 and 8), which confirms the key role played by the insured percentage of the initial investment.

$\square$

9.3.2 The dynamic behavior of OBPI and CPPI

In many situations, the use of traded options is not possible. For example, the portfolio to be insured may be a diversified fund for which no single option is available. The insurance period may also not coincide with the maturity of a listed option. Thus, for all these reasons, the OBPI put often has to be synthesized. In this framework, both CPPI and OBPI induce dynamic management of the insured portfolio. In what follows:

- First, it is proven that the OBPI method is a generalized CPPI with a variable multiple. The study of this multiple allows quantification of its risk exposure.
- Second, portfolio rebalancing implies hedging risk. Hence, hedging properties of both methods are to be analyzed, in particular the behavior of the quantity to invest on the risky asset at any time during the management period.

For the CPPI method, the key parameter is the multiple. Indeed, it determines the amount invested in the risky asset at any time. Does there exist such an "implicit" parameter for the OBPI?

9.3.2.1 OBPI as a generalized CPPI

PROPOSITION 9.8

For the geometric Brownian motion case, the OBPI method is equivalent to the CPPI method in which the multiple is allowed to vary and is given by

$$m^{OBPI}\left(t,S_t\right)=\frac{S_t N\left(d_1\left(t,S_t\right)\right)}{C(t,S_t,K)}. \tag{9.53}$$

This coefficient is the ratio of the delta of the Call $N\left(d_1\left(t,S_t\right)\right)$ multiplied by the risky asset price S_t. It is equal to the risk exposure divided by the cushion value, which is equal to the Call value $C(t,S_t,K)$.

In this framework, the OBPI method looks like a CPPI one.

Note that:

- The generalized multiple m^{OBPI}, associated to the OBPI method, is a decreasing function of the risky asset price S, at any time t.
- The multiple m^{OBPI} takes higher values than usual CPPI multiples, except when the associated Call is in-the-money.
- This implies that, for a bullish market, the OBPI method limits more the risk exposure.

The following two figures illustrate these properties.

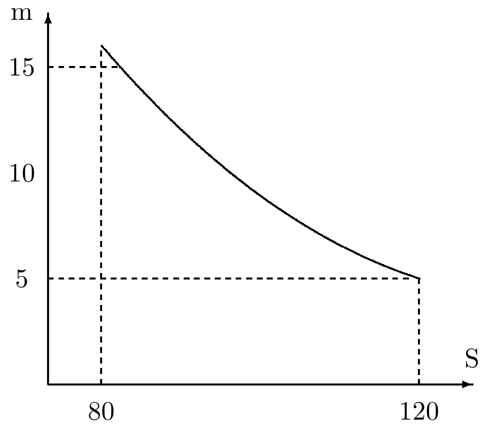


FIGURE 9.12: Multiple OBPI as function of S

The OBPI multiple takes higher values than the standard CPPI multiple, except when the associated call is in-the-money. In particular, in a rising market, the OBPI method prevents the portfolio being over-invested in the risky asset, as the multiple is low.

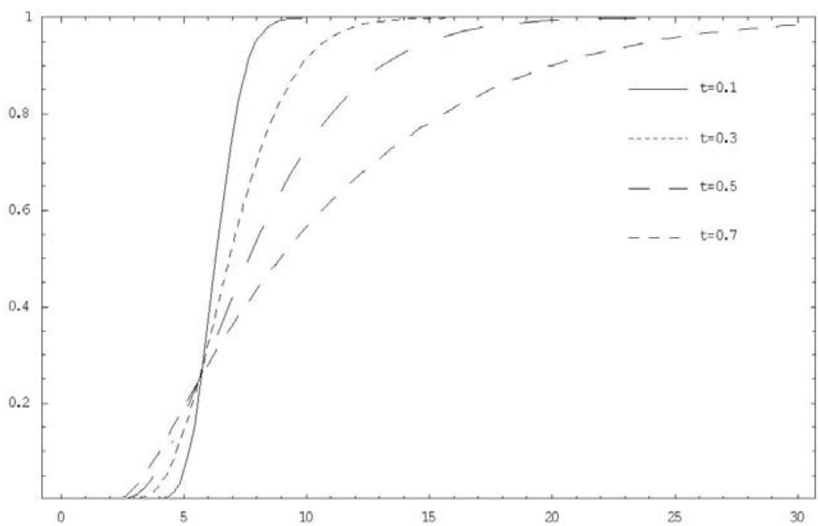


FIGURE 9.13: OBPI multiple cumulative distribution

We now study the dynamic properties of the two strategies, and in particular their "greeks."

9.3.2.2 The Delta

The delta of the OBPI is obviously the delta of the call. For the CPPI, it is given by:

$$\Delta^{CPPI} = \frac{\partial V_t^{CPPI}}{\partial S_t} = \alpha m S_t^{m-1}. \tag{9.54}$$

The following figure shows the evolution of the delta as a function of the risky asset value S_t.

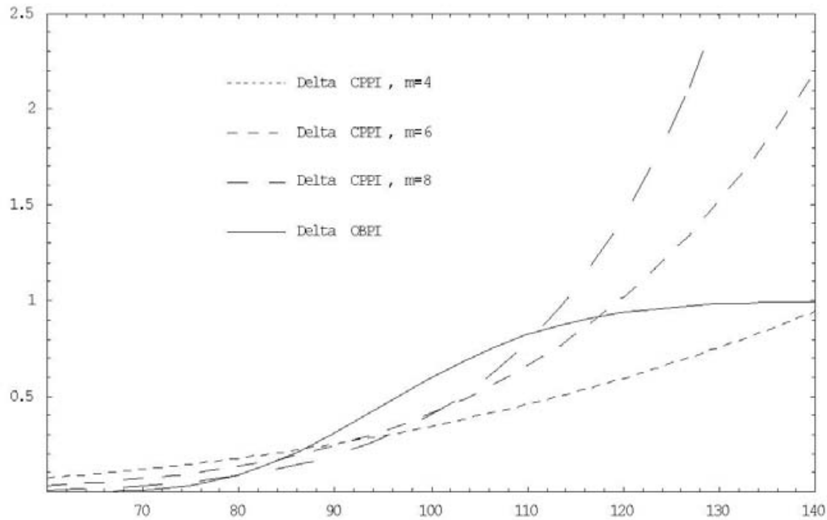


FIGURE 9.14: CPPI and OBPI delta as functions of S

It can be observed in the previous figure that the behavior of the delta of the two strategies are different. For the CPPI, not surprisingly, the delta becomes more convex with m and the delta can be greater than one.

For a large range of the values of the risky asset, the delta of the OBPI is greater than that of the CPPI. Moreover, this happens for the most likely values of the underlying asset (*i.e.*, around-the-money).

In order to be more precise, the probability that the delta of the OBPI is greater than that of the CPPI has to be calculated for various market parametrizations.
It can be observed that, in probability, CPPI is significantly less sensitive to the risky asset than OBPI, as shown in the following tables. Notice that this finding has important practical implications.

TABLE 9.3: Probability $P[\Delta^{OBPI} > \Delta^{CPPI}]$ for different m and σ

m	$\sigma = 5\%$	$\sigma = 10\%$	$\sigma = 15\%$	$\sigma = 20\%$	$\sigma = 25\%$
3	1.000	0.991	0.970	0.945	0.921
4	1.000	0.987	0.961	0.930	0.876
5	1.000	0.983	0.946	0.860	0.759
6	0.999	0.978	0.884	0.748	0.672
7	0.999	0.949	0.782	0.661	0.636
8	0.999	0.881	0.685	0.616	0.630
9	0.992	0.788	0.616	0.599	0.640
10	0.96	0.69	0.58	0.60	0.66

TABLE 9.4: Probability $P[\Delta^{OBPI} > \Delta^{CPPI}]$ for different m and μ

m	$\mu = 5\%$	$\mu = 10\%$	$\mu = 15\%$	$\mu = 20\%$	$\mu = 25\%$
3	0.925	0.930	0.943	0.951	0.953
5	0.861	0.860	0.850	0.830	0.801
6	0.774	0.748	0.712	0.667	0.616
7	0.706	0.661	0.610	0.554	0.495
8	0.670	0.616	0.558	0.497	0.436
9	0.657	0.599	0.537	0.475	0.413
10	0.657	0.598	0.536	0.473	0.411

The previous features are made clearer by examining the distribution of the ratio $\frac{\Delta^{OBPI}}{\Delta^{CPPI}}$. The next figure shows that the probability that the CPPI delta is smaller than the OBPI one is a decreasing function of the strike K (or equivalently, of the insured amount). Note that, for small values of K, the range of possible values of the ratio $\frac{\Delta_t^{OBPI}}{\Delta_t^{CPPI}}$ spreads out.

The following figure shows the evolution of the delta with time. Whatever the level of S compared to the level of the insured level at maturity, K, the delta of the CPPI is decreasing with time. For the OBPI, the evolution of the delta obviously depends on the moneyness of the option.

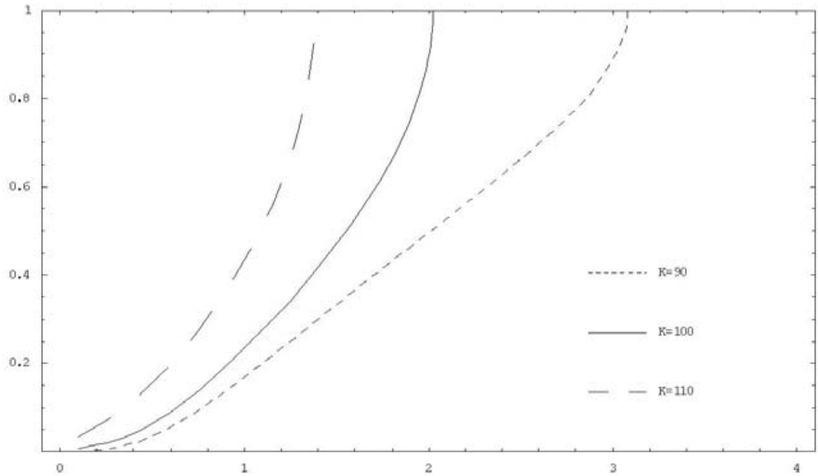


FIGURE 9.15: Cumulative distribution of $\frac{\Delta_t^{OBPI}}{\Delta_t^{CPPI}}$ at $t = 0.5$

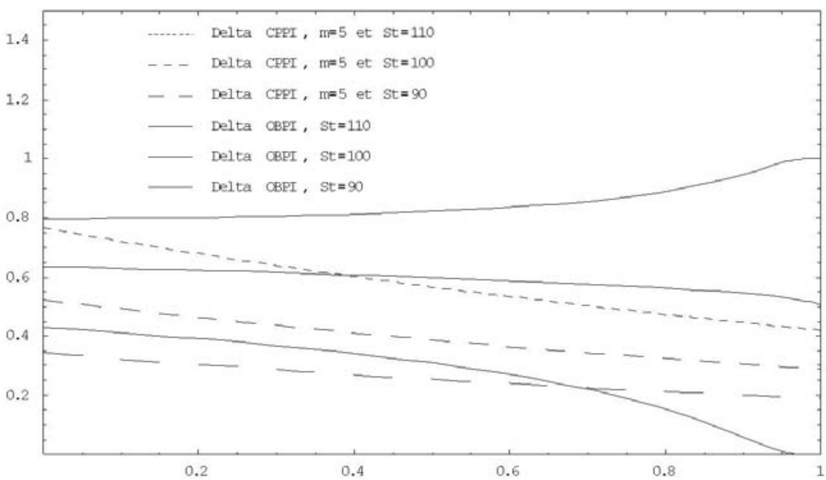


FIGURE 9.16: CPPI and OBPI delta as functions of current time.

More surprisingly, the delta of the CPPI is decreasing with the actual volatility (since $m > 1$). The same feature arises when examining the vega of the CPPI, since they depend in the same way on this actual volatility. For the OBPI, the result depends on the moneyness of the option.

9.3.2.3 The Gamma

The gamma of the CPPI is equal to :

$$\Gamma^{CPPI} = \frac{\partial \Delta_t^{CPPI}}{\partial S_t} = \alpha m(m-1) S_t^{m-2}$$

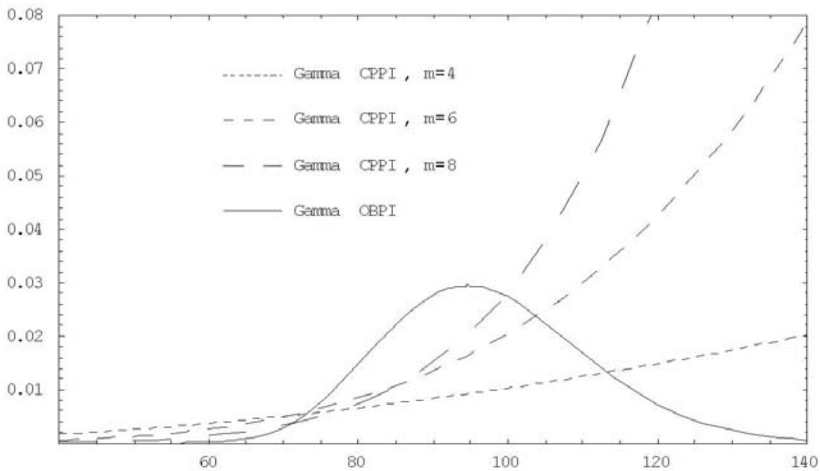


FIGURE 9.17: CPPI and OBPI gamma as functions of S at $t = 0.5$, for $K = 100$

For the CPPI, it is always for high values of S that the gamma is important. Nevertheless, for usual values of m, the CPPI gamma is smaller than the OBPI one for a large range of values of S. This is particularly true for $K = 110$.

This fact is important, as the magnitude of transaction costs are directly linked to the gamma. Again, the CPPI method seems to be better suited when the insured percentage, p, of the initial investment is high.

Moreover, the gamma of the CPPI is monotonically decreasing with time, although it does not reach zero at maturity. Recall that, for a call, the gamma will go to zero as the expiration date approaches if the call is in-the-money or out-of-the-money, but will become very large if it is exactly at-the-money.

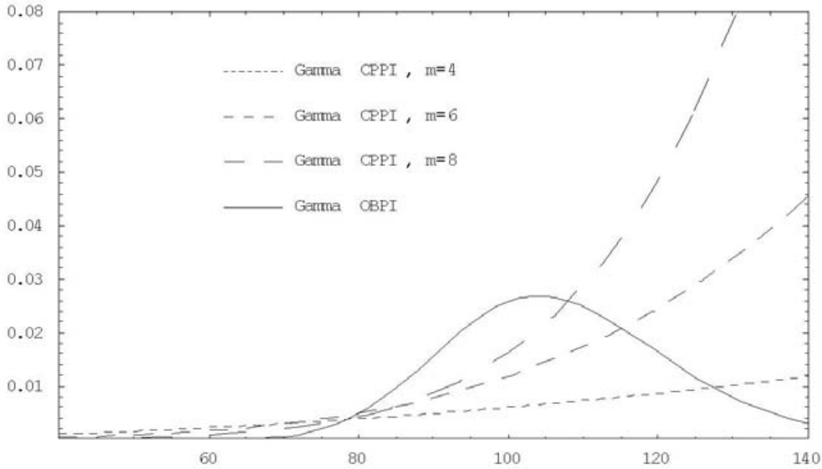


FIGURE 9.18: CPPI and OBPI gamma as functions of S at $t = 0.5$, for $K = 110$

9.3.2.4 The Vega

The vega of the CPPI is defined as[1]:

$$\begin{aligned} vega^{CPPI} &= \frac{\partial V_t^{CPPI}}{\partial \sigma} \\ &= C(0, S_0, K) \left(\frac{S_t}{S_0}\right)^m \left((m - m^2)\sigma t\right) \exp\left[\beta t\right] \\ &= \left((m - m^2)\sigma t\right) V_t^{CPPI} \end{aligned}$$

Thus, the sensitivity of the CPPI value with respect to the actual volatility is negative if $m > 1$.

The higher the multiple, the more it decreases.

REMARK 9.10 To summarize, comparison of OBPI and CPPI, with usual criteria such as first order stochastic dominance and various moments of their rates of return, does not allow one to discriminate clearly between the two strategies:

- The standard CPPI method is based on dynamic portfolio management and thus seems more flexible.

[1]In the following calculation, we do not take into account the effect of the volatility on $C(0, S_0, K)$ because the call enters in the CPPI formula only to insure the compatibility with the OBPI at time 0. Furthermore, $C(0, S_0, K)$ depends only on the expected volatility and not on the actual one.

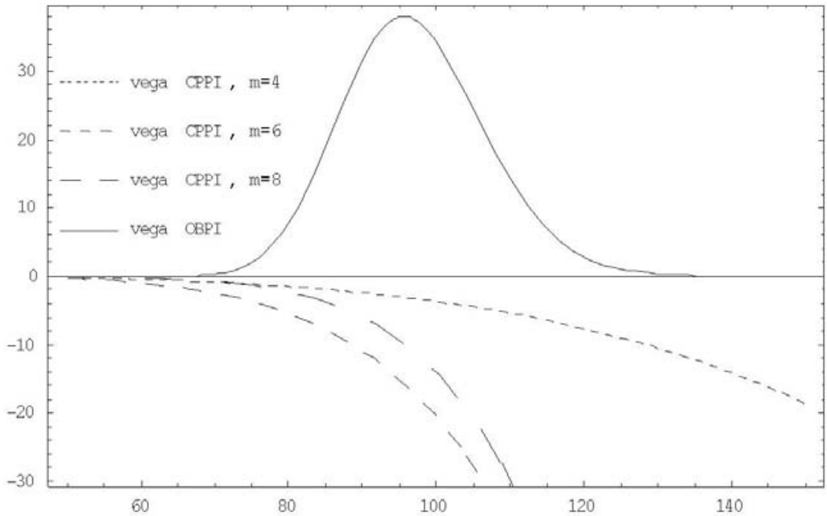


FIGURE 9.19: CPPI and OBPI vega as functions of S at $t = 0.5$

- However, to avoid sudden drops, the multiple must be not too high. In such case, the OBPI strategy seems more robust if the option is well hedged. The worst scenario for the CPPI strategy compared to the OBBI one is a sudden drop during the management period, then a financial market rise. In that case, for the standard method, the cushion remains null and it is no longer possible to benefit from risky asset price increases.

- For small risky asset values, the CPPI method provides higher returns but, in that case, a simple riskless investment can beat the CPPI portfolio.

- As the guaranteed percentage increases, the CPPI strategy is more relevant than the OBPI one. This arises mainly because the OBPI call has less chance to be exercised.

- The analysis of the dynamic properties of these two methods shows in particular how the OBPI method can be considered as a generalized CPPI method. They differ mainly by their vegas.

Note that implicit volatilities can be considered to better analyze the influence of such a parameter since the Call option is priced from implicit volatility. For the CPPI method, a conditional multiple can be considered, based on factors such as the realized or implicit volatilities.

$\square$

9.4 Further reading

As mentioned in Leland and Rubinstein (1988), long term returns can benefit from portfolio insurance methods since they allow investment in more agressive securities, while satisfying guarantee constraints.

Bookstaber and Langsam [82] analyze properties of portfolio insurance models. They focus on path dependence, showing that only option-replicating strategies provide path independence. They deal also with the problem of the time horizon and, in particular, time-invariant or perpetual strategies (studied also in Black and Perold [77]).

Black and Rouhani [75] compare CPPI with OBPI when the put option has to be synthesized. They compare the two payoffs and examine the role of both expected and actual volatilities. They show that "OBPI performs better if the market increases moderately. CPPI does better if the market drops or increases by a small or large amount."

Bertrand and Prigent [58] introduce general marked point processes to model stock price variations. Upper bounds are provided in such framework, using VaR type criteria. Such an approach can be further extended with other risk measures.

Time varying multiples can be introduced according to market turbulence factors, as proposed by Hamidi *et al.* [280] using quantile regression to estimate the potential losses, conditionally to these factors.

Cesari and Cremonini [110] examine various dynamic asset allocations by using Monte Carlo simulations. Risk adjusted performance measures, such as Sharpe ratio, Sortino ratio, and return to risk are introduced, and strategies are compared for different market evolutions (bullish, bearish, and no-trend). CPPI strategies seems to perform better in bear and no-trend markets. Obviously, benchmarking strategies are better in bullish markets, but such strategies provide no true "absolute" insurance against market drops, since they "only" must be close to the benchmark (relative performance versus absolute performance).

Chapter 10

Optimal dynamic portfolio with risk limits

Portfolio insurance payoff is designed to give the investor the ability to limit downside risk while allowing some participation in upside markets. Such methods allow investors to recover, at maturity, a given percentage of their initial capital, in particular in falling markets.

This payoff is a function of the value at maturity of some specified portfolio of common assets, usually called the benchmark. As is well-known by practitioners, specific insurance constraints on the horizon wealth must be generally satisfied. For example, a minimum level of wealth and some participation in the potential gains of the benchmark can be guaranteed. However, institutional investors for instance may require more complicated insurance contracts.

As seen previously in Chapter 9, two standard portfolio insurance methods are the *Option Based Portfolio Insurance* (OBPI) and the *Constant Proportion Portfolio Insurance* (CPPI).

However, to what extent are these methods optimal?

The literature on portfolio optimality generally considers an investor who maximizes the expected utility of his terminal wealth by trading in continuous time, as seen in Part 3. The continuous-time setup is also usually introduced to study portfolio insurance (see e.g., Grossman and Vila [274], Basak [46], or Grossman and Zhou [276]). El Karoui, Jeanblanc and Lacoste [192] prove that, *under a fixed guarantee at maturity*, an option based portfolio strategy is optimal for quite general utility functions (see also Jensen and Sorensen [302] for a particular case and [415] for a more general guarantee). The key assumption is that markets are complete which means that all portfolio profiles at maturity can be perfectly hedged.

Other literature is devoted to the optimal positioning problem which has been addressed in the partial equilibrium context by Brennan and Solanki [88] and by Leland [347]. The value of the portfolio is a function of the benchmark in a one period set up. An optimal payoff, maximizing the expected utility,

is derived. It is shown that it depends crucially on the risk aversion of the investor. Following this approach, Carr and Madan [106] consider markets in which exist out-of-the-money European puts and calls of all strikes. As they mentioned, this assumption allows the examination of the optimal positioning in a complete market and is the counterpart of the assumption of continuous trading. This approximation is justified when there is a large number of option strikes (*e.g.* for the S&P500). Due to practical constraints, liquidity, transaction costs, *etc.*, portfolios are in fact discretely rebalanced.

In this chapter, the optimal insured portfolio is determined for the two cases:

- In the first section, the insurance is perfect, since the probability that the portfolio value is above the guaranteed level is equal to 1.
  - First, a one-period market is considered. Structured portfolios with payoffs defined as functions of the risky asset (a financial stock index for example) are examined. Constraints on the horizon wealth are included. In addition, markets can be incomplete. The insured optimal portfolio is characterized for arbitrary utility functions, return distributions, and for any choice of a particular risk neutral probability if the market is incomplete.
  - Second, the financial market is assumed to be dynamically complete. In this framework, results concerning European guarantees are extended to more general guarantee constraints at maturity. For example, this guarantee can no longer be a fixed percentage of the initial investment, but can involve a stochastic component.
- In the second section, the insurance is satisfied at a given probability level.
  - First, the maximization of the probability of success is studied.
  - Second, the expected utility maximization under VaR/CVaR constraints is examined.

In particular, the optimal portfolio is calculated for CRRA utility functions.

10.1 Optimal insured portfolio: discrete-time case

10.1.1 Optimal insured portfolio with a fixed number of assets

In this section, we consider an investor who chooses a buy-and-hold strategy and can invest in three available financial assets:

- A riskless asset B;
- A risky asset S; and,
- A put written on S with strike K and initial value $P_0(K)$.

Note that this portfolio strategy is an extension of the OBPI method for which the number of shares for the underlying asset S and the Put written on S are equal.

The investor's strategy consists of setting constant porfolio shares where:

α is the number of shares invested in B;
β is the number of shares invested in S; and,
γ is the number of shares invested in the Put with strike K.

Denote by V_T the portfolio value:

$$V_T = \alpha B_T + \beta S_T + \gamma (K - S_T)^+. \tag{10.1}$$

The budget constraint is given by:

$$V_0 = \alpha B_0 + \beta S_0 + \gamma P_0(K). \tag{10.2}$$

Assume that the guarantee constraint at maturity is linear:

$$V_T \geq a S_T + b,$$

where a corresponds to a minimal percentage of the potential rise of the risky asset, and b is a fixed insured amount.

Three main cases have to be considered.

Case 1: $0 < \beta < \gamma$ (the number of purchased puts is higher than the number of purchased risky assets S).

This asset allocation allows provision of the guarantee $\alpha B_T + \beta K$. Note that the portfolio profile is convex. However, this function is decreasing for small S values, which is not satisfactory for most investors.

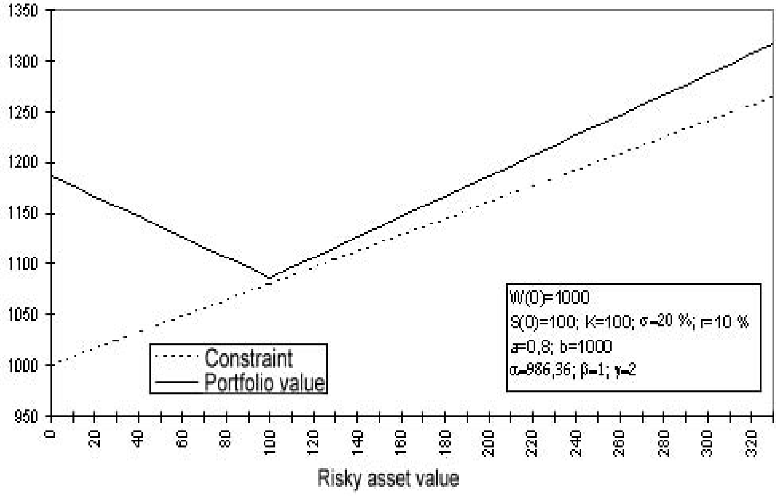


FIGURE 10.1: Optimal portfolio profiles according to instantaneous stock return (convex but not increasing)

Case 2: $\beta > \gamma > 0$ (the number of purchased puts is smaller than the number of purchased risky assets S).

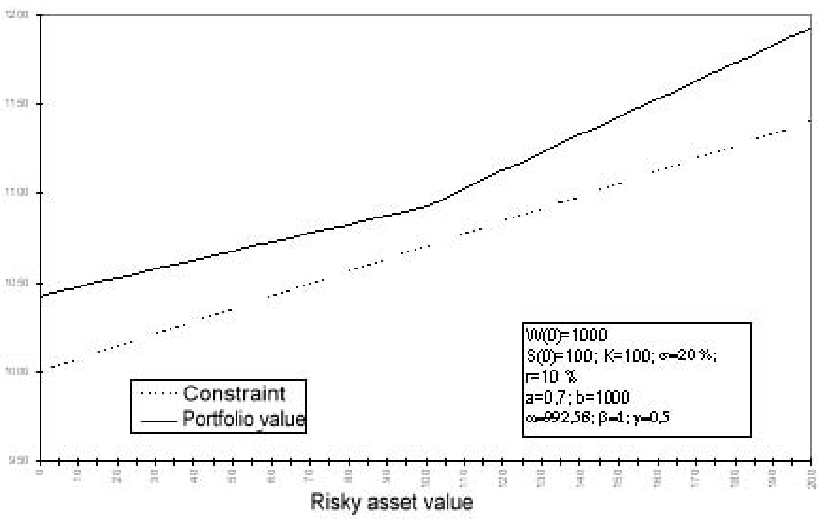


FIGURE 10.2: Optimal portfolio profiles according to instantaneous stock return (convex increasing)

In this case, the investor has a guarantee equal to $\alpha B_T + \gamma K$. The portfolio payoff is still convex.

Case 3: $\beta > 0 > \gamma$ (the investor buys S and sells the Put)

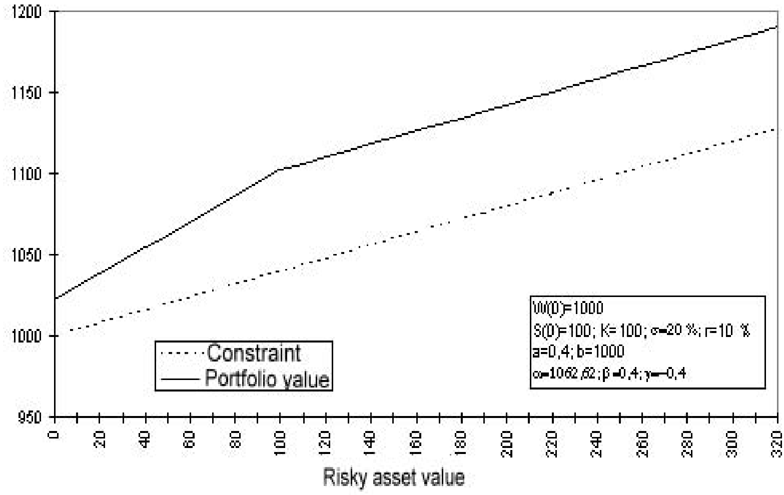


FIGURE 10.3: Optimal portfolio profiles according to instantaneous stock return (concave)

The guarantee is still equal to $\alpha B_T + \gamma K$. However, the portfolio profile is now concave. The investor will receive more money when the risky asset decreases, but less money when it increases.

We now search the optimal portfolio for an investor who maximizes his utility function U, which is assumed to satisfy usual assumptions.

REMARK 10.1 More generally, the guarantee constraints are infinite-dimensional since they must be satisfied for all S values (for example, if the probability distribution of S is lognormal). However, since both portfolio payoffs and guarantee constraints are piecewise linear, the guarantee constraints are reduced only to three conditions: the portfolio values must be above the guarantee function for $S = 0$ and $S = K$, and the slope for high values of S must be higher than the slope of the guarantee function.

$\square$

Therefore, the optimization problem is the following:

$$
\begin{array}{lll}
 & \text{Max} & \mathbb{E}[U(V_T)] \\
\text{with} & \begin{array}{l}\text{(guarantee} \\ \text{constraint)}\end{array} & \left\{\begin{array}{l} \alpha B_T + \gamma K \geq b \\ \alpha B_T + \beta K \geq aK + b \\ \beta \geq a \end{array}\right. , \\
\text{and} & \begin{array}{l}\text{(budget} \\ \text{constraint)}\end{array} & V_0 = \alpha B_0 + \beta S_0 + \gamma P_0(K).
\end{array}
$$

The solution is determined by using the Kuhn and Tucker approach. The Lagrangian is defined by:

$$
\begin{aligned}
\mathcal{L} = & \, E[U(V_T)] + \lambda_1(\alpha B_T + \gamma K - b) + \lambda_2(\alpha B_T + \beta K - aK - b) \\
& + \lambda_3(\beta - a) + \lambda_4(V_0 - \alpha B_0 - \beta S_0 - \gamma P_0(K)),
\end{aligned}
$$

where λ_4 is the Lagrange multiplier associated to the budget constraint, and where parameters λ_i are associated to guarantee constraints. They are all positive and satisfy the necessary conditions of local optimality given by:

$$
\begin{aligned}
\tfrac{\partial \mathcal{L}}{\partial \alpha} &= \mathbb{E}[U'(V_T)] + \lambda_1 + \lambda_2 - \lambda_4 e^{-rT} && = 0, \\
\tfrac{\partial \mathcal{L}}{\partial \beta} &= \mathbb{E}[U'(V_T)S_T] + \lambda_2 K + \lambda_3 - \lambda_4 S_0 && = 0, \\
\tfrac{\partial \mathcal{L}}{\partial \gamma} &= \mathbb{E}[U'(V_T)(K - S_T)^+] + \lambda_1 K - \lambda_4 P_0(K) && = 0.
\end{aligned}
$$

Therefore, eight cases have to be considered according to the three guarantee constraints which can be (or not) saturated. Three main results are observed:

Strategy 1: The investor is weakly risk-averse.

Strategy 2: The investor is risk-neutral.

Strategy 3: The investor is rather strongly risk-averse.

The weakly risk-averse investor chooses a portfolio which is equal to the guarantee for values of the risky asset S smaller than the strike K. The risk-averse investor has higher returns for small risky asset values, but smaller returns for higher values of S. According to risk aversion, the porfolio profile is concave or convex.

Assume for example that the utility function is quadratic:

$$U(x) = x - \frac{a}{2}x^2, \text{ with } a > 0.$$

Then, the number of shares can be examined according to the aversion to the variance a:

The higher the aversion to the variance a, the higher the amount invested in the riskless asset, and the smaller both the numbers of shares invested in the risky asset and in the put.

These properties are illustrated by the following figures.

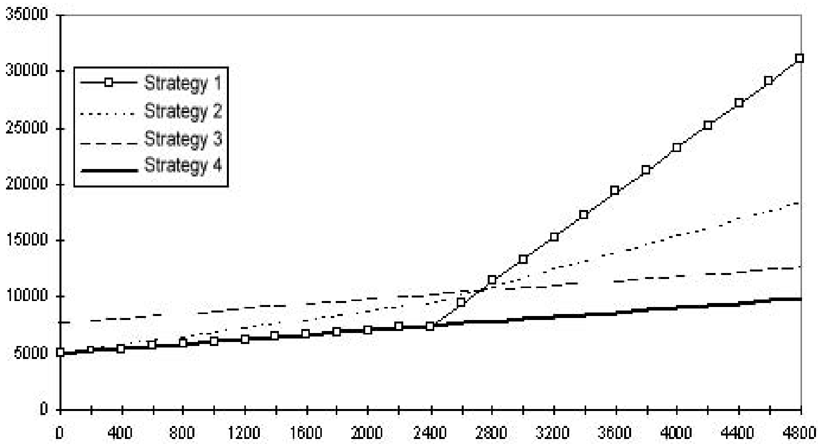


FIGURE 10.4: Optimal portfolio profile as a function of the risky asset value

$$B_0 = 150,\ S_0 = 2500,\ V_0 = 10000,\ K = 2400,\ r = 4\%,\ \sigma = 20\%,\ \mu = 5\%.$$

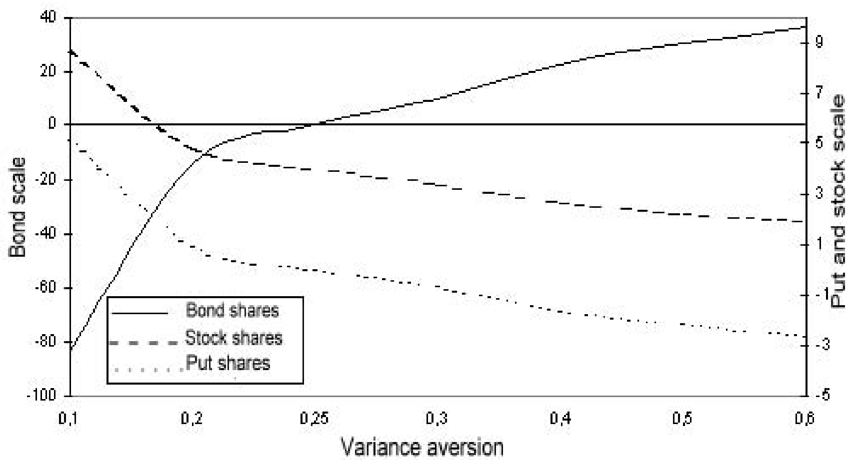


FIGURE 10.5: Optimal portfolio weighting as a function of the variance aversion (quadratic case)

10.1.2 Optimal insured payoffs as functions of a benchmark

In this section, the financial market is to be composed of three basic financial assets: the cash associated to a discount factor N, the bond B, and the stock S (a financial index, for example). We suppose that the investor determines an optimal payoff h which is a function defined on all possible values of the assets (N, B, S) at maturity. As in Section 7.1, when there is no insurance constraint, the optimization problem is solved as follows:

Assume that prices are determined under measure $\mathbb{Q}$. Denote by $\frac{d\mathbb{Q}}{d\mathbb{P}}$ the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to the historical probability P. Denote by N_T the discount factor, and by M_T the product $N_T \frac{d\mathbb{Q}}{d\mathbb{P}}$.

The investor has to solve the following optimization problem:

$$Max_h \mathbb{E}_{\mathbb{P}}[U(h(N_T, B_T, S_T)] \text{ under } V_0 = \mathbb{E}_{\mathbb{P}}[h(N_T, B_T, S_T)M_T]. \tag{10.3}$$

Assume that $h \in \mathbb{L}^2(\mathbb{R}^{+3}, \mathbb{P}_{X_T}(dx))$ where $X_T = (N_T, B_T, S_T)$ which is the set of the measurable functions with squares that are integrable on $\mathbb{R}^{+3}$ with respect to the distribution $\mathbb{P}_{X_T}(dx)$.

Now, the investor introduces a specific guarantee, which can be institutional, or may imply an additional insurance against risk. If, for example, the interest rate is not stochastic, such a guarantee can be modelled by letting a function h_0 be defined on the possible values of the benchmark S_T: whatever the value of S_T, the investor wants to get a final portfolio value above the floor $h_0(S_T)$. For instance, if h_0 is linear with $h_0(s) = as + b$, then, when the benchmark falls, the investor is sure of getting at least b (equal to a fixed percentage of his initial investment), and if the benchmark rises, he make profits out of the rises at a percentage a.

The optimal payoff with insurance constraints on the terminal wealth is the solution of the following problem:

$$\begin{gathered} Max_h \mathbb{E}_{\mathbb{P}}[U(h(X_T)] \\ V_0 = \mathbb{E}_{\mathbb{P}}[h(X_T)M_T], \\ h(X_T) \geq h_0(X_T). \end{gathered}$$

As can be seen, the initial investment V_0 must be higher than $\mathbb{E}_{\mathbb{P}}[h_0(X_T)M_T]$ if the insurance constraint must be satisfied.

The solution of this problem is given in Prigent [415]. To solve it, introduce the sets

$$H_1 = \{h \in \mathbb{L}^2(\mathbb{R}^{+3}, \mathbb{P}_{X_T}) | V_0 = \mathbb{E}_{\mathbb{P}}[h(X_T)M_T]\},$$

and

$$H_2 = \{h \in \mathbb{L}^2(\mathbb{R}^{+3}, \mathbb{P}_{X_T}) | h \geq h_0\}.$$

The set $H = H_1 \cap H_2$ is a convex set of $\mathbb{L}^2(\mathbb{R}^{+3}, \mathbb{P}_{X_T})$. Consider the following indicator function of H, denoted by δ_H and defined by:

$$\delta_H(h) = \begin{cases} 0 & if\ h \in H, \\ +\infty & if\ h \notin H. \end{cases}$$

Since H is closed and convex, δ_H is lower semi-continuous and convex.

Recall the notion of subdifferentiability (see, *e.g.*, Ekeland and Turnbull [187] for the definition and properties of subdifferentials).

Let V denote a Banach space and $< .,. >$ the duality symbol.

DEFINITION 10.1 *1) For any function F defined on V with values in $I\!R \cup \{+\infty\}$, a continuous affine functional $l : V \to I\!R$ everywhere less than F. This means that:*

$$\forall v \in V,\ l(v) \leq F(v) \text{ is exact at } v^* \text{ if } l(v^*) = F(v^*). \tag{10.4}$$

2) A function $F : V \to I\!R \cup \{+\infty\}$ is subdifferentiable at v^* if there exists a continuous affine functional $l(.) =< ., v_l > -a$, everywhere less than F, which is exact at v^*. The slope v_l of such an l is a subgradient of F at v^*. The set of all subgradients of F at v^* is the subdifferential of F at v^* and is denoted by $\partial F(v^*)$.

Recall the following characterization:

$$v_c \in \partial F(v^*) \text{ iff } F(v^*) < +\infty \text{ and } \forall v \in V, < v - v^*, v_c > +F(v^*) \leq F(v). \tag{10.5}$$

Denote by $\partial \delta_H$ the subdifferential of δ_H. The optimization problem is equivalent to:

$$Max_h \left(\mathbb{E}[U(h(X_T)] - \delta_K(h) \right). \tag{10.6}$$

The optimality conditions leads to:

PROPOSITION 10.1

There exists a scalar λ_c and a function h_c defined on $L^2(I\!R^{+3}, P_{X_T})$ such that:

$$h^* = J(\lambda_c g + h_c), \tag{10.7}$$

where λ_c is the solution of:

$$y?, V_0 = \int_0^\infty J[yg(x) + h_c(x)]g(x)f(x)dx, \tag{10.8}$$

and $h_c \in \partial \delta_{H_2}(h^*)$.

To explain more precisely the condition $h_c \in \partial \delta_{H_2}(h^*)$, assume that the functions h_0 and h_c are continuous (such properties are always verified in practice). Consequently, the optimal payoff h^* is continuous.

COROLLARY 10.1

Under the above assumption, the function h_c satisfies the following property:
1) If on a product I of intervals of values of X_T, $h^*(X_T) > h_0(X_T)$ then h_c is equal to 0 on I.
2) If on a product I of intervals of values of X_T, $h^*(X_T) = h_0(X_T)$ then h_c is negative on I.

REMARK 10.2 From the previous results, the optimal payoff h^* can be determined by introducing the unconstrained optimal payoff h^e associated to the modified coefficient λ_c (*i.e.*, $h^e = J(\lambda_c g)$). λ_c can also be considered as a Lagrange multiplier associated to a non insured optimal portfolio but with a modified initial wealth. Indeed, when h^e is greater than the insurance floor h_0, then $h^* = h^e$. Otherwise, $h^* = h_0$. However, the payoff is usually a continuous function of the values of the benchmark like any linear combination of standard options. In that case, the optimal payoff is given by:

$$h^* = Max(h_0, h^e) = h_0 + Max(h^e - h_0, 0). \tag{10.9}$$

Consequently, the optimal solution is the sum of the constraint h_0 with a call of "strike" h_0 and underlying h^e, which is the optimal solution without constraint. This result is true even in the non-complete case, and with a general guarantee constraint. In the next section ("dynamic" case), the same kind of result is proved (see Proposition 10.3). ▯

Since the general result in the previous proposition is an extension of the Kuhn-Tucker theorem to infinite dimension, it is well-known that the determination of h^* implies the comparison of all possible solutions of the kind (10.9). However, this problem can easily be solved if the payoff must be continuous.

COROLLARY 10.2

Under the previous assumptions on the utility U, there is one and only one continuous optimal payoff, associated to the unique solution λ_c of the budget equation.

PROOF From the assumptions on the marginal utility U', we deduce that its inverse J is a continuous and decreasing function with:

$$lim_{o^+} J = +\infty \text{ and } lim_{+\infty} J = 0.$$

Thus, for all s, the function $\lambda_c \longrightarrow h^*(\lambda_c, x) = Max(h_0(x), \tilde{h}(\lambda_c, x))$ is continuous and decreasing. Therefore, the function $\lambda_c \longrightarrow \mathbb{E}_{\mathbb{Q}}[h^*(\lambda_c, X_T)]$ is continuous and decreasing from $+\infty$ to $\mathbb{E}_{\mathbb{Q}}[h_0(\lambda_c, X_T)]$, which is lower than the initial investment V_0. From the intermediate values theorem and by monotonicity, the result is deduced. ▯

REMARK 10.3 Finally, an investment strategy is associated to the optimal payoff, which can be computed by the following approach. Suppose for example that the interest rate is non-stochastic. Then:

- First, the function h^* is approximated by a sequence of twice differentiable payoff functions h_n.
- Second, as proved in Carr and Madan (1997), it is possible to explicitly identify the position that must be taken in order to achieve a given payoff h_n that is twice differentiable.

h_n is duplicated by an unique initial position of $h_n(S_0) - h'_n(S_0)S_0$ unit discount bonds, $h'_n(S_0)$ shares and $h_n(K)dK$ out-of-the-money options of all strikes K:

$$\begin{aligned} h_n(S) = {} & [h_n(S_0) - h'_n(S_0)S_0] + h'_n(S_0)S \\ & + \int_0^{S_0} h''_n(K)(K-S)^+ dK + \int_{S_0}^{\infty} h''_n(K)(S-K)^+ dK. \end{aligned}$$

□

Generally, as mentioned previously, h_0 is increasing and h^e also. Therefore, the optimal payoff is an increasing function of the benchmark.

REMARK 10.4 From the previous theoretical result, we conclude that:
- Generally, an optimal portfolio must include options in order to maximize the expected utility of investors.
- The solution is a combination of the optimal portfolio value without guarantee, and a put written on it with a strike equal to the floor. Under the standard assumptions that the insurance constraints and the payoff are modelled by continuous functions of the risky asset, the solution is also the maximum between this function and the solution of the unconstrained problem but with a different initial wealth.
- In the no guarantee case, the concavity/convexity of the portfolio profile is determined from the degree of risk aversion and from the financial market performance, for example a Sharpe-type ratio. This kind of result still holds according to the insurance constraint at maturity.

□

All the above properties are illustrated in the next example.

10.1.2.0.1 A special case

Assume that the utility function of the investor is a CRRA utility:

$$U(x) = \frac{x^\alpha}{\alpha},$$

with $0 < \alpha < 1$, from which we deduce $J(x) = x^{\frac{1}{\alpha-1}}$.

Suppose that the interest rate r is constant, that the stock price evolves in a continuous time set up, and in particular that $(S_t)_t$ is a geometric Brownian motion given by:

$$S_t = S_0 exp\left[(\mu - 1/2\sigma^2)t + \sigma W_t\right].$$

Denote by f the density of S_T.

Notations:

$$\theta = \frac{\mu - r}{\sigma},\ A = -\frac{1}{2}\theta^2 T + \frac{\theta}{\sigma}(\mu - \frac{1}{2}\sigma^2)T,$$
$$\psi = e^A (S_0)^{\frac{\theta}{\sigma}},\ \kappa = \frac{\theta}{\sigma}.$$

Recall that in the Black and Scholes model, the conditional expectation g of $\frac{d\mathbb{Q}}{d\mathbb{P}}$ under the σ-algebra generated by S_T is given by:

$$g(s) = \psi s^{-\kappa}.$$

Therefore, $h^e(s)$ satisfies:

$$h^e(s) = d \times s^m \text{ with } \quad d = c\psi^{\frac{1}{\alpha-1}} \text{ and } \ m = \frac{\kappa}{1-\alpha} > 0. \tag{10.10}$$

We apply the previous general results to solve the optimization problem. Then, if there is no insurance constraint, the optimal payoff is given by:

$$h^e(s) = \frac{V_0 e^{rT}}{\int_0^\infty g(s)^{\frac{\alpha}{\alpha-1}} f(s)ds} \times g(s)^{\frac{1}{\alpha-1}}. \tag{10.11}$$

If the insurance constraint is required then the optimal payoff must be the solution of

$$\begin{array}{c} Max_h \mathbb{E}[\frac{(h(S_T))^\alpha}{\alpha}] \\ V_0 = e^{-rT}\mathbb{E}[h(S_T)] \\ h(S_T) \geq h_0(S_T) \end{array}$$

Then:

PROPOSITION 10.2

The optimal payoff with guarantee is given by:

$$h^* = (\lambda_c g + h_c)^{\frac{1}{\alpha-1}}, \tag{10.12}$$

where λ_c is the solution of:

$$y?, V_0 e^{rT} = \int_0^{\infty} [yg(s) + h_c(s)]^{\frac{1}{\alpha-1}} g(s) f(s) ds, \tag{10.13}$$

and h_c is a negative function satisfying the property of the previous corollary.

COROLLARY 10.3
Assume as usual that h_0 is increasing and continuous. Recall that we have $h^* = Max(h^e, h_0)$ and the solution h^e of the unconstrained problem associated to the Lagrange multiplier λ_c is increasing. Then, the optimal payoff is an increasing continuous function of the benchmark at maturity.

$h^* = Max(h^e, h_0)$ and the solution h^e of the unconstrained problem associated to the Lagrange multiplier λ_c is increasing.

REMARK 10.5 As seen in Chapter 7, if there is no insurance constraint, the concavity/convexity of the optimal payoff is determined by the comparison between the risk-aversion and the ratio $\kappa = \frac{\mu - r}{\sigma^2}$, which is the Sharpe ratio divided by the volatility σ.

i) h^e is concave if $\kappa < 1 - \alpha$.
ii) h^e is linear if $\kappa = 1 - \alpha$.
iii) h^e is convex if $\kappa > 1 - \alpha$.

The graph of the optimal payoff changes from concavity to convexity according to the increase of the risk-aversion of the investor. If, for example, the insurance constraint is linear ($h_0(s) = as + b$), it looks like the unconstrained case, except when h^* is equal to the constraint h_0.

□

The previous theoretical result justifies the introduction of power options in the portfolio (see Chapter 9).

As seen in Section 9.2, if $\kappa < 1 - \alpha$, h^{**} the optimal payoff is concave, and if $\kappa > 1 - \alpha$, h^{**}, it is convex.

As shown previously for the buy-and-hold case (one bond, one stock, and only a finite number of options written on it), if the guaranteed payoff is linear, the optimal (polygonal) payoff is still convave/convex according to the degree of risk aversion.

Example 10.1
Consider the same parameter values as in Example 9.2.

Then, the optimal payoff profile can be examined according to values of parameters d and m. Assuming that

$$\mu = 0.1, \sigma = 0.2, V_0 = 100, S_0 = 100, r = 3, T = 5, a = 0.7, b = 90, \alpha = 0.7,$$

the optimal payoff profile is equal to:

$$h^*(s) = d.s^m,$$

with $d = 8.7253.10^{-11}$ and $m = 5.8333$.

The terminal portfolio value is given by:

$$V_T^{**} = 0.7S_T + 90 + \max\left(8.7253.10^{-11}.S_T^{5.8333} - 0.7S_T - 90; 0\right).$$

Note for example that:

$$\alpha = 0.1 \Rightarrow d = 0.01 \text{ and } m = 0.84,$$
$$\alpha = 0.9 \Rightarrow d = 1.55.10^{-36} \text{ and } m = 17.5.$$

□

REMARK 10.6 When the number of available options is finite, the indirect expected utility is smaller than the previous one, where an infinite number of options can be used to replicate the portfolio value $h^{**}(s)$.

This static replication is based on the following result (see, *e.g.*, Carr and Madan [106]):

Any payoff $p(s)$ can be replicated by a position which is composed of

- $[p(S_0) - (\partial p/\partial s)(S_0)S_0]$ shares of riskless asset;
- $[(\partial p/\partial s)(S_0)]$ shares of risky asset S; and,
- $[\left(\partial^2 p/\partial s^2\right)(K)]$ shares of out-of-the money options for all strikes K.

However, for a given finite set of available options, we can consider either a combination $p(s)$ of these options which minimizes a given loss function w.r.t. the optimal payoff $h^{**}(s)$, or the optimal solution for the given utility function with the finite set of available options.

□

10.2 Optimal Insured Portfolio: the dynamically complete case

10.2.1 Guarantee at maturity

Assume that the financial market is complete, arbitrage free, and frictionless. Asset prices are supposed to follow continuous time diffusion processes. According to the investor's risk aversion and horizon, the portfolio manager chooses the proportions to invest on financial assets, among them all zero-coupon bonds for maturities T defined on an instantaneous interest rate $(r_t)_t$. The resulting portfolio value $(V_t)_t$ is self-financing. This means that the process $(V_t \exp(-\int_0^t r_s ds))_t$ is a $\mathbb{Q}$-martingale where $\mathbb{Q}$ is the risk-neutral probability.

Denote by $\eta = \frac{d\mathbb{Q}}{d\mathbb{P}}$ the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to the historical probability $\mathbb{P}$. Denote also by M_T the process $\eta_T \exp(-\int_0^T r_s ds)$. Due to the no-arbitrage condition, the budget constraint corresponds to the following relation:

$$V_0 = \mathbb{E}_{\mathbb{Q}}[V_T \exp(-\int_0^T r_s ds)] = \mathbb{E}_{\mathbb{P}}[V_T M_T].$$

Assume that the investor wants to maximize an expected utility under the statistical probability $\mathbb{P}$. As usual, the utility U of the investor is supposed to be increasing, concave, and twice-differentiable. Suppose also that the marginal utility U' satisfies:

$$\lim_{o^+} U' = +\infty \text{ and } \lim_{+\infty} U' = 0.$$

Denote by J the inverse of the marginal utility U'.

The guarantee constraint consists in letting the portfolio value V_T at maturity above a floor F_T. This floor may be deterministic, corresponding, for example, to a predetermined percentage p of the initial invesment V_0, or may be stochastic if, for instance, the investor wants to benefit from potential market rises. For example, this floor may be equal to

$$F_T = aS_T + b,$$

where a is a given percentage of the benchmark S (a stock index, for instance) and b is a fixed guaranteed amount which corresponds usually to a fixed percentage of the initial investment. In all cases, it is assumed that there exists a portfolio that duplicates the floor F_T.

Then, for a given initial investment V_0^*, the investor wants to find the portfolio θ solution of the following optimization problem:

$$\text{Max}_\theta \ \mathbb{E}_{\mathbb{P}}[U(V_T)] \text{ under } V_T \geq F_T.$$

Due to market completeness, this problem is equivalent to (see Cox-Huang [132]):

$$\begin{aligned} &\mathrm{Max}_{V_T}\ \mathbb{E}_{\mathbb{P}}[U(V_T)] \\ &\text{under } V_T \geq F_T \text{ and } V_0^* = \mathbb{E}_{\mathbb{P}}[V_T M_T] \geq \mathbb{E}_{\mathbb{P}}[F_T M_T]. \end{aligned} \tag{10.14}$$

PROPOSITION 10.3

The optimal solution V_T^* of problem (10.14) is given by the maximum of the floor F_T and the solution V_T^e of the non constrained problem for an initial investment V_0^e such that $V_0^* = \mathbb{E}_{\mathbb{P}}[Max(V_T^e, F_T)M_T]$. Equivalently, this solution can be viewed as a combination of the portfolio value V_T^e and a put written on it with "strike" equal to the floor, or a combination of the floor and a call written on the portfolio value V_T^e.

$$V_T^* = V_T^e + (F_T - V_T^e)^+ = F_T + (V_T^e - F_T)^+.$$

PROOF Consider the solution V_T^* of the free problem (without guarantee constraint). Using Cox and Huang [132] results, this solution is given by:

$$V_T^e = J(\alpha M_T),$$

where the Lagrangian parameter α is such that $V_0^e = \mathbb{E}_{\mathbb{P}}[V_T^e M_T]$.

Furthermore, for any portfolio V_T with initial investment V_0 satisfying $V_T \geq F_T$, since the marginal utility U' is concave, we have:

$$U(V_T) - U(V_T^*) \leq U'(V_T^*)(V_T - V_T^*),$$

and since U' is decreasing, we deduce:

$$U'(V_T^*)(V_T - V_T^*) = \mathrm{Min}(\alpha M_T, U'(F_T))(V_T - V_T^*).$$

Additionally,

$$\mathrm{Min}(\alpha M_T, U'(F_T))(V_T - V_T^*) = \alpha M_T(V_T - V_T^*) - [\alpha M_T - U'(F_T)]^+(V_T - F_T).$$

Finally, since $\mathbb{E}_{\mathbb{P}}[V_T M_T] = V_0^* = \mathbb{E}_{\mathbb{P}}[V_T^* M_T]$, we get:

$$\mathbb{E}_{\mathbb{P}}[\mathrm{Min}(\alpha M_T, U'(F_T))(V_T - V_T^*)] = -\mathbb{E}_{\mathbb{P}}[[\alpha M_T - U'(F_T)]^+(V_T - F_T)] \leq 0.$$

Therefore:

$$\mathbb{E}_{\mathbb{P}}[U(V_T)] \leq \mathbb{E}_{\mathbb{P}}[U(V_T^*)].$$

$\square$

10.2.2 Risk exposure and utility function

Call-power options and CPPI strategy can be chosen according to optimality criteria, as seen in the next examples. In what follows, we assume that the interest rate r is constant and that $(S_t)_t$ is a geometric Brownian motion given by:

$$S_t = S_0 exp\left[(\mu - 1/2\sigma^2)t + \sigma W_t\right].$$

As done previously, we denote:

$$\theta = \frac{\mu - r}{\sigma},\ A = -\frac{1}{2}\theta^2 T + \frac{\theta}{\sigma}(\mu - \frac{1}{2}\sigma^2)T,$$

$$\psi = e^A (S_0)^{\frac{\theta}{\sigma}},\ \kappa = \frac{\theta}{\sigma}.$$

Recall that in the Black and Scholes model, the conditional expectation g of $\frac{d\mathbb{Q}}{d\mathbb{P}}$ under the σ-algebra generated by S_T is given by:

$$g(s) = \psi s^{-\kappa}.$$

Example 10.2
Assume that the investor has an HARA utility U given by:

$$U(x) = \frac{(x-K)^\alpha}{\alpha}.$$

Then, the portfolio value V_T which maximizes the expected utility is given by: there exists a non-negative constant ζ such that

$$V_T(S_T) = K + \zeta(S_T^{\frac{\kappa}{1-\alpha}}).$$

Thus, it corresponds to a CPPI portfolio value with guarantee K and a multiple equal to $\frac{\kappa}{1-\alpha}$. ∎

Example 10.3
Assume that the investor has a CRRA utility U given by:

$$U(x) = \frac{x^\alpha}{\alpha}.$$

Then, the portfolio value V_T which maximizes the expected utility with the additional guarantee constraint K at maturity is associated to a non-negative constant ξ such that:

$$V_T(S_T) = K + (\xi S_T^{\frac{\kappa}{1-\alpha}} - K)^+.$$

Thus, it corresponds to a call-power portfolio value with guarantee K and a power equal to $\frac{\kappa}{1-\alpha}$. ∎

Example 10.4
Assume that the investor has a CRRA utility U given by:

$$U(x) = \frac{x^{\alpha}}{\alpha}.$$

Then, the portfolio value V_T which maximizes the expected utility defined on the cushion $V_T - K$ is given by: there exists a non-negative constant χ such that

$$V_T(S_T) = K + (\chi S_T^{\frac{\kappa}{1-\alpha}}).$$

Thus, it corresponds also to a CPPI portfolio value with guarantee K and a multiple equal to $\frac{\kappa}{1-\alpha}$.

□

REMARK 10.7
- For a given profile $h(S_T)$, under some mild assumptions, a utility function U can be identified (up to a linear transformation) such that the solution of the expected utility maximization is the profile $h(S_T)$.

Assume for instance that h is defined and invertible on a given interval of $\mathbb{R}$. Then, the condition: there exists a non-negative scalar λ such that

$$U'(x) = \lambda g(h^{-1}(x))$$

insures that $h(S_T)$ is the optimal profile for the expected U maximization for some initial amount V_0.

- For standard options (generally not invertible), U must be determined for each case. For instance, if

$$V_T = V_0 + (S_T - K)^+,$$

then we can consider the utility function:

$$U(x) = (-\infty)\mathbb{I}_{x \leq V_0} + \frac{x - V_0 + K}{m+1} g(x - V_0 + K)\mathbb{I}_{x > V_0}.$$

Therefore, the choice of a particular insured portfolio can determine an implicit utility function.

□

10.2.3 Optimal portfolio with controlled drawdowns

As in Grossman and Zhou [275], consider an investor who wants at any time to lose no more than a fixed percentage of the maximum value his portfolio has achieved up to that time.

Denote by M_t the maximum value of the investor's wealth V invested on the portfolio, on or before time t. Thus the constraint on the value process V is as follows: there exists a constant λ in $[0,1]$ such that at any time t in $[0,T]$,

$$M_t \geq \lambda V_t.$$

- *The financial market* contains a riskless asset with rate r. The price of the risky asset S is the solution of the following SDE w.r.t. the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$:

$$dS_t = S_t.\left[\mu dt + \sigma dW_t\right], \tag{10.15}$$

where W is a standard Brownian motion, and the usual assumptions on coefficient μ and σ are made. Denote $\widetilde{\mu} = \mu - r$. This financial market is complete.

- *The portfolio value* V is the solution of:

$$dV_t = V_t\left[rdt + w_S\left(\widetilde{\mu}dt + \sigma dW_t\right)\right],$$

where w_S is the portfolio weight invested on the stock S which is assumed to be predictable w.r.t. the filtration $(\mathcal{F}_t)_t$ and such that the wealth process V is always positive.

- *The drawdown control:* Denote M_0 as a positive amount invested at time 0 which evolves at a growth rate δ with $\delta \leq r$. Define:

$$M_t = \max\left(M_0 e^{\delta t}, V_s e^{\delta(t-s)}, s \leq t\right). \tag{10.16}$$

Note that, when $\delta = 0$, M_t denotes the highest value between the initial value M_0 and all the portolio values on or before time t. If $\delta \to -\infty$, M_t goes to 0.

The portfolio value must always be above the stochastic floor λM :

$$\forall t \in [0,T],\ V_t \geq \lambda M_t,\ a.s. \tag{10.17}$$

The term $\left(1 - \frac{V_t}{M_t}\right)$ is called the drawdown. Therefore, the drawdown control condition is: for a given level λ,

$$\left(1 - \frac{V_t}{M_t}\right) \geq 1 - \lambda. \tag{10.18}$$

- *The utility function:* Assume that the investor has a power utility function $U(x) = \frac{x^\alpha}{\alpha}$ with $\alpha < 1$ and $\alpha \neq 0$. Suppose that the objective is to maximize

the long-term growth rate of the expected utility. Its upper value is given by:

$$\xi^* = \sup_{w_S \in \mathcal{A}} \lim \inf_{T \to \infty} \frac{1}{\alpha T} \ln \left(\mathbb{E}\left[\alpha U(V_T) \right] \right), \tag{10.19}$$

where $\mathcal{A}$ is the set of weights w_S such that condition 10.17 is satisfied.

Recall that, when $\lambda = 0$, the optimal investment strategy is given by:

$$w_S^* = \frac{1}{1-\alpha} \frac{\mu - r}{\sigma^2} \text{ and } \xi^* = r + \frac{1}{2(1-\alpha)} \frac{(\mu - r)^2}{\sigma^2}.$$

Note that when $\lambda > 0$, this strategy violates condition (10.17) with probability one.

Denote $\widetilde{M_t} = e^{-\delta t} M_t$ and $\widetilde{V_t} = e^{-\delta t} V_t$.

PROPOSITION 10.4 *Grossman and Zhou [275]*
If there exists a constant ξ and a function $\mathcal{J}(\widetilde{V}, \widetilde{M})$ such that:
i) $\mathcal{J}(\widetilde{V}, \widetilde{M})$ is solution of the Bellman equation:

$$\mathcal{J}(\widetilde{V}, \widetilde{M})_t = \sup_{w_S \in \mathcal{A}} \mathbb{E}\left[\alpha \mathcal{J}(\widetilde{V}_t, \widetilde{M}_t) U(V_T) e^{-\alpha \xi t} \right].$$

ii) There exists a trading strategy w_S^* which achieves the supremum; and,
iii) There exist positive constants C_1 and C_2 such that:

$$C_1 \alpha U(V) \leq \alpha \mathcal{J}(\widetilde{V}, \widetilde{M}) \leq C_2 \alpha U(V),$$

then, the maximum long-term growth rate of the expected utility of final wealth is achieved by the strategy w_S^*. In addition, ξ is also the rate of the finite-horizon problem:

$$\xi = \lim \inf_{T \to \infty} \frac{1}{\alpha T} \ln \left(\alpha \sup_{w_S \in \mathcal{A}} \mathbb{E}\left[U(V_T) \right] \right), \tag{10.20}$$

Moreover, consider $\overline{\mathcal{J}}(\widetilde{V}, \widetilde{M})$ defined by:

$$\overline{\mathcal{J}}(\widetilde{V}, \widetilde{M}) = \sup_{w_S \in \mathcal{A}} \lim \inf_{T \to \infty} \mathbb{E}\left[U(V_T) e^{-\alpha \xi T} \right].$$

Then, if $\overline{\mathcal{J}}(\widetilde{V}, \widetilde{M})$ is finite for $\widetilde{V} \geq \lambda \widetilde{M}$, we have:
- The functional $\overline{\mathcal{J}}(\widetilde{V}, \widetilde{M})$ is homogeneous of degree α in $\widetilde{V}$ and $\widetilde{M}$.
- $\overline{\mathcal{J}}(\widetilde{V}, \widetilde{M})$ is an increasing and concave function of $\widetilde{V}$ and a decreasing function of $\widetilde{M}$.

Example 10.5 *Case* $\lambda = 0$ *and* $\delta = r$

- The optimal weight on stock is given by:

$$w_{S,t}^* = k\frac{V_t - \lambda M_t}{V_t} = k\left(1 - \lambda\frac{M_t}{V_t}\right),$$
$$\text{with } k = \frac{\mu - r}{\sigma^2}\frac{1}{(1-\lambda)(1-\alpha)+\lambda}.$$

When $\lambda = 0$, we recover the Merton's solution.

Thus, for CRRA utility functions, the optimal strategy is equivalent to an investment in the risky asset which is proportional to the cushion $V_t - \lambda M_t$. This is analogous to the CPPI method (see Chapter 9) but with a stochastic floor λM_t. When we assume $M_t = K$, we recover the standard CPPI method.

The optimal weight $w_{S,t}^*$ is an increasing function of the ratio $\frac{M_t}{V_t}$. When the portfolio value V_t is close to the floor λM_t, the optimal weight $w_{S,t}^*$ is close to 0, since the guarantee must not be violated.

- The optimal portfolio value V is the solution of the following SDE:

$$dV_t = k(V_t - \lambda M_t)\left[(\mu - r)dt + \sigma dW_t\right] \text{ with } M_t = \max(M_0, V_t).$$

The process $\ln\left(\frac{V_t}{M_t} - \lambda\right)$ is a regulated Brownian motion. Then, we deduce:

$$V_t = \lambda M_0 \exp\left((1-\lambda)L_t\right) + (V_0 - \lambda M_0)\exp\left[-\lambda L_t + \left(k\mu - \frac{1}{2}k^2\sigma^2\right)t + k\sigma W_t\right],$$

with

$$L_t = \max_{s\leq t}\left[0\ ;\ \ln\left(\frac{V_0}{M_0} - \lambda\right) + \left(k\mu - \frac{1}{2}k^2\sigma^2\right)s + k\sigma W_s - \ln(1-\lambda)\right].$$

$\square$

10.3 Value-at-Risk and expected shortfall based management

In this section, the guarantee condition is satisfied at a given probability threshold, not neccessarily equal to 1. Two cases are examined:

- First, the analysis of portfolio asset management under dynamic safety constraints which control the probability that the yield falls under a given level.

- Second, the expected utility maximization under VaR/CVaR constraints.

10.3.1 Dynamic safety criteria

Consider safety criteria such as those of Roy [437] and Kataoka [324], which are presented in a one-period setting, in Chapter 3. In what follows, we use results from Prigent and Toumi [418] and Toumi [492]. To simplify the exposition, the financial market is assumed to evolve in continuous time and is supposed to be dynamically complete (see [492] for the the discrete-time and incomplete cases).

10.3.1.1 Roy Criterion

There exists a riskless asset taken as numeraire. The risky asset price S is supposed to be a semimartingale $S = (S_t)_{t\in[0,T]}$ on a complete probability space $(\Omega, \mathcal{F}, P)$ with a filtration $(\mathcal{F}_t)_{t\in[0,T]}$.

The predictable strategy process ξ is assumed to be self-financing with an initial invested amount $V_0 > 0$. Such strategy (V_0, ξ) is called admissible if the process defined by:

$$V_t = V_0 + \int_0^t \xi_s dS_s, \forall t \in [0,T], \mathbb{P} - a.s. \tag{10.21}$$

satisfies:

$$V_t \geq 0 \qquad \forall t \in [0,T], \mathbb{P} - a.s.$$

Since the financial market is assumed to be complete, there exists a unique equivalent martingale measure $\mathbb{Q} \approx P$.

According to Roy's criterion, the investor searchs for the best admissible strategy (V_0,ξ) which maximizes the probability that the portfolio terminal value verifies $V_T \geq e^{rT} R_{Min} V_0$, for a given minimal fixed return R_{Min}.

Therefore, the investor has to solve the following optimization problem:

$$\max_{\xi} \mathbb{P}\left[V_0 + \int_0^T \xi_s dS_s \geq R_{Min} V_0\right]. \tag{10.22}$$

For fixed R_{Min}, consider the "success set" $A_{(V_0,\xi)}$ defined by:

$$A_{V_0,\xi} = \{V_T \geq R_{Min} V_0\}. \tag{10.23}$$

First step: We show that the Roy problem is equivalent to the determination of a success set of maximal probability:

PROPOSITION 10.5

Let $\widetilde{A} \in \mathcal{F}_T$ be a solution to the problem

$$\max \mathbb{P}[A], \tag{10.24}$$

under the constraint:

$$\mathbb{E}_{\mathbb{Q}}\left[\mathbb{I}_A\right] \leq \frac{1}{R_{Min}}. \tag{10.25}$$

Let $\widetilde{\zeta}$ be the perfect hedge of the option $\widetilde{H} = \mathbb{I}_{\widetilde{A}} \in \mathbb{L}^1(\mathbb{Q})$, i.e,

$$\mathbb{E}_{\mathbb{Q}}\left[\mathbb{I}_{\widetilde{A}}/\mathcal{F}_t\right] = \mathbb{E}_{\mathbb{Q}}\left[\mathbb{I}_{\widetilde{A}}\right] + \int_0^t \widetilde{\zeta}_s dS_s \qquad \forall t \in [0,T], \qquad \mathbb{P}-a.s. \tag{10.26}$$

Then $\left(V_0; \xi = \frac{V_0}{\mathbb{E}_{\mathbb{Q}}[\mathbb{I}_{\widetilde{A}}]}\widetilde{\xi}\right)$ is the solution of problem (10.22), and the corresponding success set $A_{\tilde{V}_0,\tilde{\xi}}$ is equal almost surely to $\widetilde{A}$.

REMARK 10.8 - The proof is detailed in [492]. It is based on results of quantile hedging, provided in Föllmer and Leukert [234]. An alternative proof is given in Browne [95] for the (Gaussian) complete case.

- The problem of constructing a maximal success set according to Roy criteria is then solved by applying Neyman-Pearson lemma, in a similar manner to the case of quantile theory in Föllmer and Leukert [234]. In fact, the Roy problem can be viewed as a quantile hedging of the constant $H = R_{Min} V_0$ under the constraint that the initial capital V_0 to invest is smaller than H and that R_{Min} is higher than 1 (note that here the riskless return is equal to 1). □

Since obviously $\mathbb{Q}[A] = \mathbb{E}[\mathbb{I}_A]$, problem (10.24) is equivalent to the maximization of $\mathbb{P}[A]$ under the constraint $\mathbb{Q}[A] \leq \frac{1}{R_{Min}}$. This is the reason why we can apply the Neyman-Pearson lemma.

Let $\tilde{a}$ be the threshold defined by:

$$\widetilde{a} = \inf\left\{a : \mathbb{Q}[\frac{d\mathbb{P}}{d\mathbb{Q}} > aH] \leq \frac{1}{R_{Min}}\right\}, \tag{10.27}$$

and consider the set

$$\widetilde{A} \leq \left\{ \frac{d\mathbb{P}}{d\mathbb{Q}} > \tilde{a} \right\}. \tag{10.28}$$

PROPOSITION 10.6
Let's assume that the set $\widetilde{A}$ defined by the two previous relations satisfies:

$$\mathbb{Q}[\tilde{A}] = \tilde{a}. \tag{10.29}$$

Then the optimal strategy $\left(V_0; \xi = \frac{V_0}{\mathbb{E}_{\mathbb{Q}}[\mathbb{I}_{\widetilde{A}}]}\widetilde{\xi}\right)$ is the solution to problem (10.22).

Second step: We maximize the expected success ratio.

The condition $\mathbb{Q}[\tilde{A}] = \tilde{a}$ is clearly satisfied when:

$$\mathbb{P}\left[\frac{d\mathbb{Q}}{d\mathbb{Q}} = \widetilde{a}\right] = 0. \tag{10.30}$$

Generally, it is not easy to find a set $\tilde{A} \in \mathcal{F}_T$ defined by (10.30). In this case, the Neyman-Pearson theory suggests the replacement of the critical region $A \in \mathcal{F}_T$ with a random test, i.e, by a function φ, $\mathcal{F}_T$-measurable, such that $0 \leq \varphi \leq 1$.

Let $\mathcal{R}$ be the set of these functions φ and consider the following optimization problem:

$$\mathbb{E}\left[\widetilde{\varphi}\right] = \max_{\varphi \in \mathcal{R}} \mathbb{E}\left[\varphi\right] \tag{10.31}$$

under the constraint

$$\mathbb{E}_{\mathbb{Q}}\left[\widetilde{\varphi}\right] \leq \frac{1}{R_{Min}}. \tag{10.32}$$

The Neyman-Pearson lemma proves that the solution $\widetilde{\varphi}$ of (10.31, 10.32) has the form:

$$\widetilde{\varphi} = \mathbb{I}_{\left\{\frac{d\mathbb{P}}{d\mathbb{Q}} > \widetilde{a}H\right\}} + \gamma \mathbb{I}_{\left\{\frac{d\mathbb{P}}{d\mathbb{Q}} = \widetilde{a}H\right\}}, \tag{10.33}$$

where $\widetilde{a}$ is given by (10.29), and where γ is defined by:

$$\gamma = \frac{1/R_{Min} - \mathbb{Q}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} > \widetilde{a}\right]}{\mathbb{Q}\left[\frac{d\mathbb{P}}{d\mathbb{Q}} = \widetilde{a}H\right]}. \tag{10.34}$$

This allows a solution for the dynamic Roy problem:

DEFINITION 10.2 *Let (V_0, ξ) be an admissible strategy. We define the "success ratio" associated to this strategy as*

$$\varphi_{V_0,\xi} = \mathbb{I}_{\{V_T \geq R_{Min}V_0\}} + \frac{V_T}{R_{Min}V_0}\mathbb{I}_{\{V_T < R_{Min}V_0\}}. \tag{10.35}$$

Note that $\varphi_{V_0,\xi} \in \mathcal{R}$ and the set $\{\varphi_{V_0,\xi} = 1\}$ coincide with the success set $A_{V_0,\xi} = \{V_T \geq V_0 \ R_{Min}\}$ associated to the strategy (V_0, ξ).

We search for the strategy maximizing the mean of the success rate $\mathbb{E}\left[\varphi_{V_0,\xi}\right]$ under the probability $\mathbb{P}$ over the set of admissible strategies:

$$\max\left\{\mathbb{E}\left[\varphi_{V_0,\xi}\right] : (V_0, \xi) \ \text{admissible}\right\}. \tag{10.36}$$

PROPOSITION 10.7

Let $\widetilde{\xi}$ represent the process determining the perfect hedging of the $\mathcal{F}_T$−measurable function $\widetilde{\varphi}$ defined by (10.33). Then the strategy $\left(V_0; \frac{V_0}{R_{Min}} \widetilde{\xi}\right)$ is a solution to (10.36).

Example 10.6

Examine the Roy criterion in the Black and Scholes framework.

The price process S of the risky asset is given by:

$$dS_t = S_t(mdt + \sigma \ dW_t),$$

where W is a Brownian motion under $\mathbb{P}$ and m is constant. To simplify, the interest rate is assumed to be null $r = 0$.

The unique equivalent martingale measure is defined by:

$$\frac{d\mathbb{Q}_T}{d\mathbb{P}_T} = \exp\left[-\frac{m}{\sigma} W_T - \frac{1}{2}\left(\frac{m}{W}\right)^2 T\right],$$

The process $W_t^* = W_t + \frac{m}{\sigma}\ t$, is a Brownian motion under $\mathbb{Q}$ and the risky asset price is equal to:

$$S_t = S_0 \exp\left(\sigma W_T^* - \frac{1}{2}\sigma^2 T\right).$$

For fixed R_{Min}, the optimal Roy strategy is the replicating strategy of the option $\widetilde{H} = I_{\widetilde{A}}$ where $\widetilde{A}$ is written as $\widetilde{A} = \{\frac{d\mathbb{P}_T}{d\mathbb{Q}_T} > \widetilde{a}\}$, and where $\widetilde{a}$ is determined from the relation

$$\mathbb{E}_{\mathbb{Q}}[\mathbb{I}_{\widetilde{A}}] = \frac{1}{R_{Min}}. \tag{10.37}$$

Since we have:

$$\frac{d\mathbb{P}_T}{d\mathbb{Q}_T} = \beta (S_T)^{\frac{m}{\sigma^2}},$$

with

$$\beta = (1/S_0)^{\left(\frac{m}{\sigma^2}\right)} \exp\left(-\frac{1}{2}\left(\frac{m}{\sigma^2}\right) T + \left(\frac{m}{2}\right) T\right),$$

then we deduce:

$$\widetilde{A} = \{\beta(S_T)^{\frac{m}{\sigma^2}} > \tilde{a}\}.$$

We consider only the case $m > 0$ (the financial market has a positive trend). For the case $m > 0$, the success set $\widetilde{A}$ has the following form:

$$\widetilde{A} = \{S_T > c\},$$

where c is determined from the relation $\mathbb{E}_{\mathbb{Q}}[\mathbb{I}_{\tilde{A}}] = \frac{1}{R_{Min}}$.

The Roy optimal portfolio has a payoff equal to the payoff of an option that can be written as:

$$\tilde{H} = \mathbb{I}_{\tilde{A}} = \mathbb{I}_{\{\ S_T\ >\ c\}},$$

and the Roy success probability is then given by:

$$\mathbb{P}(A) = \Phi(-\frac{b - \frac{m}{\sigma}T}{\sqrt{T}}),$$

where Φ is the cdf of the standard Gaussian distribution, and b is such that:

$$c = S_0 \exp(\sigma b - \frac{1}{2}\sigma^2 T).$$

Setting

$$d_-(c,t) = -\frac{1}{\sigma\sqrt{T}} Ln(\frac{c}{S_t}) - \frac{1}{2}\sigma\sqrt{T-t},$$

the value V_t of the option to duplicate is equal to:

$$V_t = E_t^*[I_{\tilde{A}}] = \Phi(d_-(c,t))$$

and

$$b = -\sqrt{T}\Phi^{-1}(\frac{1}{R_{Min}}).$$

The expressions of the Delta $\Delta(t, S_t) = \dfrac{\partial V_t}{\partial S_t}(t, S_t)$ and of the Gamma $\Gamma(t, S_t) = \frac{\partial V_t}{\partial S_t^2}(t, S_t)$ allow us to study the variation of the quantity invested in the risky asset.

We obtain:

$$\theta_S^q(t) = \Delta^q(t, S_t) = \frac{1}{\sqrt{2\pi T}\sigma S_t} \exp\left(-d_-(c,t)^2/2\right),$$

$$\Gamma^q(t, S_t) = \frac{-1}{\sqrt{2\pi T}\sigma S_t^2}(1 + \frac{1}{\sqrt{T}\sigma} d_-(c,t)) \exp\left(-d_-(c,t)^2/2\right).$$

∎

Consider the following numerical values:

$$S_0 = 100, \ m = 0.03, \ \sigma = 0.2, \ T = 1, \text{ and } R_{Min} = 1.2.$$

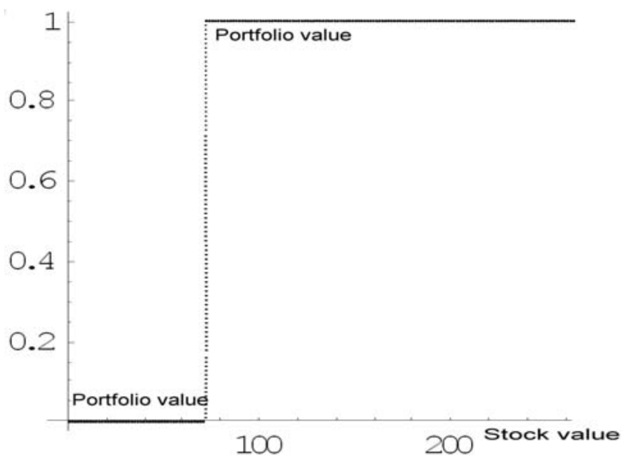


FIGURE 10.6: Dynamic Roy portfolio payoff

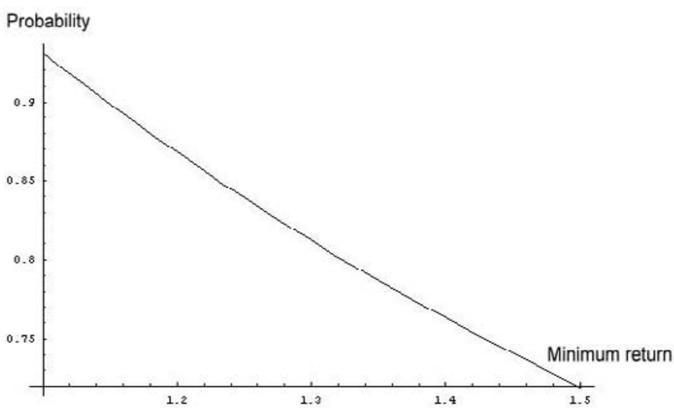


FIGURE 10.7: Probability of success as function of the minimal return R_{Min}

REMARK 10.9

- In order to maximize the success probability, the investor chooses to get exactly the minimum return for risky asset values above a given level. Indeed, if he would receive a higher return than the minimum one, the success probabililty would be reduced for the same initial investment.

- Additionally, we note that for small values of the risky asset, the optimal portfolio payoff is null, which induces extreme losses.

- To avoid such a problem, it would be preferable to use a criterion such as the expected shortfall, which takes more account of the loss sizes.

- For large values of S_t, the investor chooses to reduce the quantity $\Delta^q(t, S_t)$ invested in the risky asset. Once the condition

$$V_T \geq R_{Min} V_0$$

is achieved, the investor is satisfied and the increase of the value of the asset does not interest him any more. However, for small values of S_t, $\Delta^q(t, S_t)$ is an increasing function of S_t.

- The success probability curve is obviously a decreasing function of R_{Min}, since it is more difficult to guarantee a larger return. Besides,

$$\mathbb{P}(\widetilde{A}) = \Phi(-\frac{b - \frac{m}{\sigma}T}{\sqrt{T}})$$

is an increasing function of m.

- Note also that $\mathbb{P}(\widetilde{A})$ is a decreasing function of the volatility σ. Actually, the event $V_T \geq R_{Min} V_0$ is less probable at the expiry date if the volatility is larger.

$\square$

10.3.2 Expected utility under VaR/CVaR constraints

In what follows, we examine Value-at-Risk based management, as analyzed for example in Basak and Shapiro [47].

\- *The financial market* contains a riskless asset with rate r. The price S of the d risky assets is assumed to be a semimartingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$, and to be the solution of the SDE:

$$dS_t = diag(S_t).\left[\mu\left(t\right)dt + \sigma\left(t\right)dW_t\right], \tag{10.38}$$

where W is a $d-$dimensional standard Brownian motion and usual previous assumptions on coefficient functions are made. This financial market is complete. Thus, there exists one and only one risk-neutral probability $\mathbb{Q}$ defined as follows.

Denote the relative risk process η by:

$$\eta(t) = \sigma\left(t\right)^{-1}\left[\mu\left(t\right) - r(t)\mathbb{I}\right],$$

with:

$$\mathbb{E}\left[\int_0^T \left|\left|\eta(t)\right|\right|^2 dt\right] < \infty.$$

The exponential local martingale L:

$$L(t) = \exp\left[-\frac{1}{2}\int_0^t \left|\left|\eta(s)\right|\right|^2 ds - \int_0^t \eta(s)dW_s\right],$$

is the Radon-Nikodym density of the risk-neutral probability $\mathbb{Q}$ w.r.t. the probability $\mathbb{P}$.

Denote R the discount factor:

$$R(t) = \exp\left[-\int_0^t r(s)ds\right].$$

Denote M as the product: $M(t) = R(t)L(t)$.

\- *Expected utility maximization:* the investor has a utility function on wealth $U : (0, +\infty) \to \mathbb{R}$. Using the martingale approach, the dynamic optimization problem under a VaR constraint is the following:

$$\begin{gathered}\max \mathbb{E}\left[U(V_T)\right],\\ \text{under } \mathbb{E}\left[M_T V_T\right] \leq V_0,\\ \mathbb{P}\left[V_T \geq V_{\min}\right] \geq 1 - \varepsilon.\end{gathered}$$

PROPOSITION 10.8 *Basak and Shapiro [47]*

The optimal portfolio value V_T^* with VaR constraint is given by:

$$V_T^* = \begin{cases} J\left(yM_T\right) \text{ if } M_T < \underline{M}, \\ M_{\min} \text{ if } \underline{M} \leq M_T < \overline{M}, \\ J\left(yM_T\right) \text{ if } \overline{M} \leq M_T, \end{cases}$$

where $J = \left(U'\right)^{-1}$, M is such that $\mathbb{P}\left[M_T \geq \overline{M}\right] = \varepsilon$ and the positive Lagrange multiplier y is such that $\mathbb{E}\left[M_T V_T\right] = V_0$.

The VaR constraint is binding if and only if $\underline{M} < \overline{M}$. Also, the Lagrange parameter y is decreasing in ε and $y \in [y^B, y^{PI}]$.

Define the quantity $\underline{M_{Min}}$ by:

$$\underline{M_{Min}} = \begin{cases} J(y\overline{M}) \text{ if } \underline{M} < \overline{M}, \\ M_{Min} \text{ otherwise.} \end{cases}$$

The following figure plots the optimal portfolio value as a function of the state price M_T. The right thin curve corresponds to the portfolio B (case of no VaR contraint: $\varepsilon = 1$). The left thin curve (PI) is the portfolio value when the insurance constraint $V_T \geq M_{Min}$ must always be satisfied ($\varepsilon = 0$).

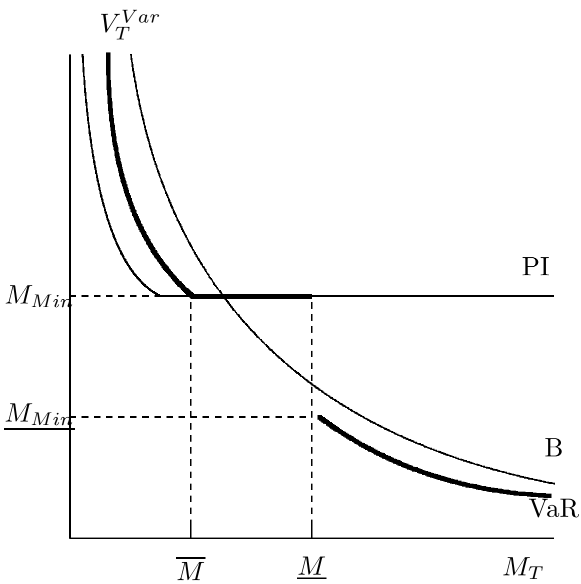


FIGURE 10.8: Optimal portfolio value with VaR constraints

REMARK 10.10 Whenever the constraint is binding, the investor must reduce portfolio losses to satisfy the VaR constraint. Then, he chooses to increase portfolios losses in the "costly" states (i.e. $M_t > \overline{M}$). Unfortunately, these events correspond to the largest losses when there is no insurance VaR constraint. Therefore, the portfolio value with a VaR constraint has a fatter left tail! ∎

PROPOSITION 10.9

(Power utility)

Assume that $U(V) = \frac{V^\alpha}{\alpha}$ *with* $\alpha < 1$ *and* r *and* η *are constant.*

- *The optimal wealth is given by:*

$$V_t^{VaR} = \frac{e^{\Gamma(t)}}{(yM_t)^{\frac{1}{1-\alpha}}}$$
$$- \left[\frac{e^{\Gamma(t)}}{(yM_t)^{\frac{1}{1-\alpha}}} \Phi(-d_1(\underline{M})) - M_{Min} e^{-r(T-t)} \Phi(-d_2(\underline{M})) \right] \tag{10.39}$$
$$+ \left[\frac{e^{\Gamma(t)}}{(yM_t)^{\frac{1}{1-\alpha}}} \Phi(-d_1(\overline{M})) - M_{Min} e^{-r(T-t)} \Phi(-d_2(\overline{M})) \right],$$

where Φ *is the cdf of the standard normal distribution and*

$$\underline{M} = \frac{1}{y\underline{M}^{1-\alpha}},$$

$$\Gamma(t) = \frac{\alpha}{1-\alpha}\left(r + \frac{||\eta||^2}{2}\right)(T-t) + \left(\frac{\alpha}{1-\alpha}\right)^2 \frac{||\eta||^2}{2}(T-t),$$

$$d_2(x) = \frac{ln(\frac{x}{M_t}) + \left(r - \frac{||\eta||^2}{2}\right)(T-t)}{||\eta||\sqrt{T-t}}$$

$$d_1(x) = d_2(x) + \frac{||\eta||\sqrt{T-t}}{1-\alpha}.$$

The fraction of wealth invested in stocks is given by:

$$w_t^{VaR} = q_t^{VaR} w_t^B,$$

where the value w_t^B *and the exposure to risky assets relative to the benchmark* B *are:*

$$w_t^B = \frac{1}{1-\alpha}[\sigma_t^{-1}]\eta,$$

$$q_t^{VaR} = 1 - \frac{\underline{\underline{M}} e^{-r(T-t)}(\Phi(-d_2(\underline{M})) - \Phi(-d_2(\overline{M})))}{V_t^{VaR}}$$
$$+ \frac{(1-\alpha)(\underline{M} - \underline{M_{Min}}) e^{-r(T-t)}(\phi(d_2(\overline{M}))}{V_t^{VaR}||\eta||\sqrt{T-t}},$$

where ϕ *is the standard normal pdf.*

10.4 Further reading

Portfolio insurance constraints of the American type are examined in El Karoui *et al.*[192]. For example, for the CRRA case, the optimal solution is based on American puts. Optimal portfolios with controlled drawdowns are also determined in Cvitanic and Karatzas [138].

The dynamic minimization of expected shortfall is detailed in Pham [409] and Föllmer and Leukert [235].

Cuoco et al. [135] introduce a dynamic VaR constraint. Contrary to the Basak and Shapiro results when the VaR constraint is static, they prove that, if the investor can fully use the current information, then the risk exposure of an investor, subject to a VaR constraint, is always lower than that of an investor with no VaR constraint.

Emmer et al. [203] also study VaR-type constraints applied to the determination of optimal portfolios with bounded Capital-at-Risk. They examine the continuous-time maximization of the expected terminal wealth under an upper bound on the Capital-at-Risk. In a Black-Scholes framework, they deduce explicit formulae. They also consider generalized inverse Gaussian diffusions for which some qualitative results can be proved and simple simulation methods can be proposed.

The equilibrium of portfolio insurance is analyzed in Grossman and Zhou [276] and Basak [48], who examine the impact of portfolio insurance methods on market volatility.

Portfolio insurance with stochastic dominance criterion is developed in El Karoui and Meziou [193] using a general concave criterion over all martingales with American constraint. The result is also used to determine utility-maximizing strategies.

Chapter 11

Hedge funds

11.1 The hedge funds industry

11.1.1 Introduction

The development of hedge funds over recent years is due to different factors:

- During the 1990s, the equity bull market had been in favor of financial investment development. Higher returns provided by more complex portfolios had earlier been requested by investors with relative weak risk-aversion and who were searching for alternative investments.

- Since the year 2000, to protect their capitals, investors have searched for hedge funds in order to diversify, and in order to limit exposure to main financial indices.

Figure 11.1 illustrates the development of the hedge funds industry.

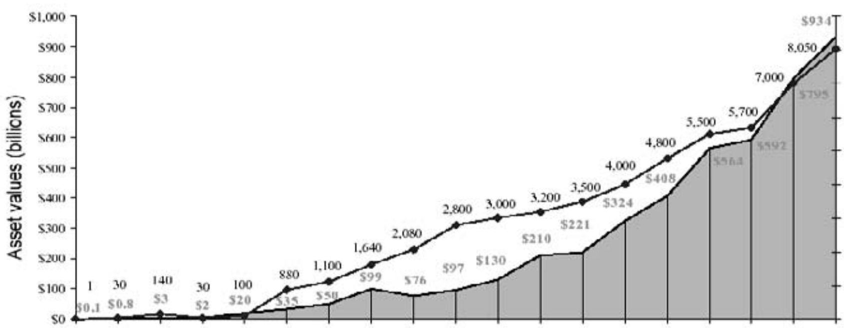


FIGURE 11.1: The hedge funds development

A hedge fund is a pool where the investor may be a (standard) shareholder with or without influence on the pool governance or may have a partnership.

11.1.2 Main strategies

The ranking of hedge funds by investment styles is not easy:

- The classification is not standardized.
- The number of subclasses is still growing.
- The relative transparency of some hedge funds does not facilitate the analysis.

Recall two usual classifications: HFR and CSFB Tremont.

TABLE 11.1: HFR and CSFB classifications

HFR	**CSFB Tremont**
Emerging markets	Emerging markets
Equity hedge	Long/Short equity
Distressed securities	
Equity market neutral	Market neutral
Equity non hedge	
Event driven	Event driven
Fixed-income	Fixed-income arbitrage
Market timing	Managed futures
Merger arbitrage	
Regulation D	
Relative value arbitrage	
Convertible arbitrage	Convertible arbitrage
Sector	
Short selling	Dedicated short bias
Statistical arbitrage	
Funds of funds	

According to Amenc *et al.* [24], the classification can be also based on principal component analysis. This method of classification includes:

- Convertible arbitrage and volatility arbitrage

  The goal is to use differences of prices for independent markets. The different risks are extracted: stock, rates, volatility, and credit. These relatively recent funds are rather complex. For example, their objective can be to provide the monetary rate plus a fixed excess return $x\%$.

- Commodities trading advisors (CTA)

  A CTA fund (for example a "Trend Follower" fund) is based on market anticipations of *futures* and *forwards*, such as the exchange markets. No correlation with usual financial markets is provided by strategies which insure diversification.

- Fusion/acquisition arbitrage

  Such funds are, for instance, the "Merger Arbitrage," "Risk Arbitrage," and "Event Driven." Simultaneously, long and short positions are taken on firms implied in merger or acquisition process.

- "Distressed" funds

  Bonds and stocks of a firm which has a bankruptcy risk will maybe have higher values in the future, if the firm can survive and grow again.

- "Long/Short Equity" funds

  This is one of the oldest alternative strategies and one of the most important, according to the amounts invested in such a way. It uses all types of assets (stocks, options, *etc.*). The idea is to shortsell midcap or bigcap stocks.

- "Fixed-Income Arbitrage" funds

  These are based on arbitrages on fixed-income markets: treasury bonds, futures and options on interest rates (*e.g.*, swaptions, caps, floors), credit derivatives, *etc.*

- "Global Macro" funds

  This strategy is based on macroeconomic variables: stock indices, exchange rates, inflation rate, taxation policy, *etc.* Using information from macro-economic factors may allow better anticipation of unusual price variations.

11.2 Hedge fund performance

11.2.1 Return distributions

Hedge fund returns are not easy to estimate, since different biases can occur, according to, for instance, Fung and Hsieh ([246],[247]):

- *The "survival" bias*;
- *The "selection" bias*; and,
- *The "instant history" bias.*

- Data are not necessarily public and, therefore, some of them are not available. A "self reporting" bias can also affect the results. The *selection bias* is also due to non-standardized selection criteria.

- The *survival bias* is due to the fact that "badly" managed funds disappear, while the other ones remain. Therefore, funds for which we have data for a relatively long period are more successful. For example, Gregoriou [268] examined data on the Zurich market along the time period from 1990-2001. The median of the life duration for a hedge fund is about 5.5 years.

- *The instant history bias* is generated by differences between dates when funds are introduced in databases. New incorporated funds have generally higher recent performances.

- Very often, the assumptions of independence and stationarity are not validated. The flexibility and the short life of such funds also induce statistical problems. Hedge fund returns are not Gaussian; for example when derivative assets are included or specific dynamic strategies are used.

Hedge funds also have risk exposures to volatility risk, default and credit risk, and liquidity problems.

Therefore, all these sources of risk must be taken into account in order to measure the performance of alternative investments.

11.2.2 Sharpe ratio limits

The performance measures introduced in Chapter 5 are mainly based on Markowitz criterion. Therefore, they are validated when standard assumptions on asset returns, such as normal distributions, are satisfied, or when investors have utility functions depending only on the first two moments.

As soon as probability distributions are no longer symmetrical, performance measures such as the Sharpe ratio may no longer be adapted. Strategies which do not require any anticipation of the fund manager can get higher Sharpe ratios than "buy and hold" strategies.
Therefore:

- We can search for alternative performance measures as in Leland [348].

- We can also profit from this inadequacy to maximize the Sharpe ratio as shown by Goetzmann *et al.*[257].

11.2.2.1 Sharpe ratio inadequacy

11.2.2.1.1 Static strategy based on options Leland [348] considers a portfolio which is long on a stock index and short on a call written on this index.

Therefore:

- The portfolio payoff is a concave function.
- This strategy consists of selling the index when it rises, and buying it when it decreases.
- Then, the portfolio skewness is reduced.
- No "*market timing*" or "*stock picking*" are required.
- Then, this type of strategy provides fairly average returns but with possible severe losses.

The following graph shows the profile of such a portfolio, for different strikes K (the call option is supposed to be priced by the Black-Scholes formula).

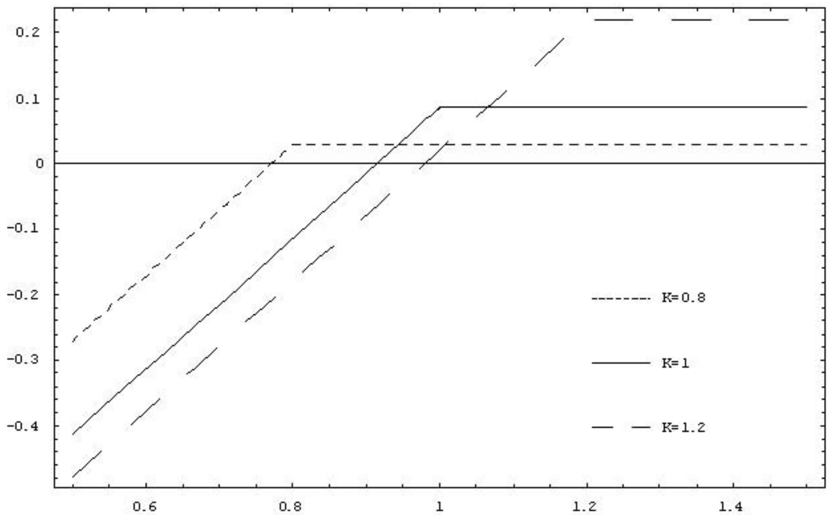


FIGURE 11.2: Portfolio profile $(S-(S-K)^{+})$ as a function of stock value S

The next table presents some of the characteristics of such a strategy for various values of the strike K.

Consider the following market parameter values:

$$S_0 = 1, \mu = 12\%, \sigma = 18\%, r = 3\%, T = 1.$$

The corresponding mean and standard deviations of the portfolio return per year are given respectively by 12.75% and 20.46%.

Notations :

$\mathbb{E}(R_P)$ denotes the mean return of portfolio P.

β_P and α_P denote the beta and Jensen alpha of portfolio P w.r.t. the index S.

RS_P is the Sharpe ratio of portfolio P.

σ_P and ε_P are respectively the total risk and specific risk of portfolio P.

TABLE 11.2: Characteristics of the portfolio $S - (S - K)^+$

	$E(R_P)$ (%)	β_P	α_P (%)	RS_P	σ_P (%)	ε_P (%)
$K = 0.9$	4.63	0.076	0.845	0.419	3.22	2.82
$K = 1$	6.30	0.199	1.329	0.482	6.18	4.65
$K = 1.1$	8.20	0.376	1.514	0.508	9.70	5.92
$K = 1.2$	9.89	0.566	1.358	0.511	13.11	6.15
$K = 1.3$	11.13	0.729	1.015	0.504	15.88	5.46
$K = 1.4$	11.92	0.846	0.658	0.495	17.84	4.31
$K = 1.5$	12.35	0.920	0.381	0.487	19.07	3.11
$\mathbf{K \to \infty}$	**12.75**	**1**	**0**	**0.474**	**20.46**	**0**

From the previous table, we can see that:

- The alpha of portfolio P is always positive, while none of these portfolios requires specific anticipation of the fund manager.

- Moreover, there exists an optimal portfolio with respect to alpha criterion. Besides, the Sharpe ratio is a function of the exercise price K.

- There exists also a value of K which maximizes the Sharpe ratio. Note that the Sharpe ratio of the index is equal to 0.4743.

- Therefore, using the Sharpe ratio, it is possible to dominate a "*buy-and-hold*" strategy by a static one based on a long position on a call, and a short position on the underlying asset.

- In addition, the total risk of the portfolio is an increasing function of the strike. At the limit, the portfolio risk converges to that of the index when the strike goes to infinity. Note also that the specific portfolio risk is first increasing then decreasing w.r.t. the strike K.

- Contrary to this strategy which always has a positive alpha, we can consider a strategy which always has a negative alpha. To do so, it is sufficient to introduce a portfolio which is made up of the index and of a put written on the index. As seen in Chapter 9, this portfolio corresponds to an insurance strategy with a payoff that is a convex function of the index.

The following figure shows the profile of such a portfolio as function of the terminal value S of the index, according to various strikes H of the put.

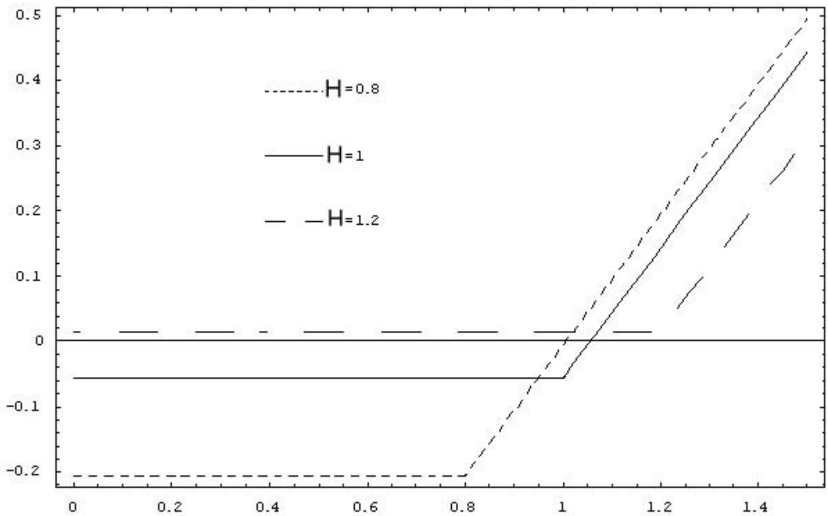


FIGURE 11.3: Portfolio profile $(S + (H - S)^{+})$

The next table provides some of the characteristics of this strategy using the same market parameters as before.

TABLE 11.3: Characteristics of the portfolio $S + (H - S)^{+}$

	Portfolio : $S + P$					
	$E(R_P)$ (%)	β_P	α_P (%)	RS_P	σ_P (%)	ε_P (%)
$\mathbf{H = 0}$	**12.75**	**1**	**0**	**0.474**	**20.46**	**0**
$H = 0.9$	11.22	0.924	−0.789	0.437	19.11	2.818
$H = 1$	9.41	0.801	−1.409	0.395	17.04	4.650
$H = 1.1$	7.34	0.624	−1.765	0.339	14.08	5.922
$H = 1.2$	5.58	0.434	−1.680	0.278	10.81	6.153
$H = 1.3$	4.38	0.271	−1.300	0.217	7.79	5.464
$H = 1.4$	3.68	0.154	−0.855	0.163	5.33	4.309

As with the previous strategy, this portfolio insurance strategy also does not require any anticipation by the fund manager. However, it generates alphas which are systematically negative and which depend on strikes.

Note that the Sharpe ratio is decreasing w.r.t. the strike K. Thus, the portfolio insurance strategy which maximizes the Sharpe ratio corresponds to the value $K = 0$. It is the index itself. Its Sharpe ratio is equal to 0.4743.

Consequently, as soon as an investment in the stock is hedged by a put written on it, its performance as measured by the Sharpe ratio is reduced and this reduction is increasing with an increasing protection (*i.e.*, the strike K increases).

However, as seen in Chapter 10, these types of strategy may be interesting for investors.

Goetzmann *et al.* [257] propose an example of a strategy that requires a perfect knowledge of the financial market, but which has a weak Sharpe ratio.

11.2.2.2 Alternative measure and optimal Sharpe ratio

Two approaches can be proposed:

- First, as proposed by Leland (1999), we can search for a modified CAPM with a beta that can better measure the portfolio risks for any probability distribution. Then, the alpha of strategies based on options would be equal to 0.

- Second, it is also possible to determine the portfolio strategy which maximizes the Sharpe ratio, as in Goetzmann *et al.* [257].

11.2.2.2.1 Alternative definition of alpha and beta for the CAPM
Leland [348] introduces an alternative definition for parameters alpha and beta in the CAPM framework. This new valuation model takes account of all the moments of the probability distribution.

It is based on a model introduced by Rubinstein [438], which is based on a valuation model through a power utility.

Under such assumptions, the CAPM is:

$$\mathbb{E}[R_P] = R_f + B_P (\mathbb{E}[R_M] - R_f), \tag{11.1}$$

where

$$B_P = \frac{Cov[R_P, -(1+R_M)^{-\gamma}]}{Cov[R_M, -(1+R_M)^{-\gamma}]}. \tag{11.2}$$

Note that for $\gamma = -1$, the utility function is quadratic. Then, we recover the standard CAPM formula, since:

$$B_P = \frac{Cov[R_P, R_M]}{Cov[R_M, R_M]}. \tag{11.3}$$

The risk aversion coefficient γ of the representative investor is given by:

$$\gamma = -\frac{\ln(\mathbb{E}[1+R_M]) - \ln(\mathbb{E}[1+R_f])}{Var[\ln(1+R_M)]}. \tag{11.4}$$

The new definition of alpha is:

$$A_P = \left(\mathbb{E}\left[R_P | I_G\right] - R_f\right) - B_P\left(\mathbb{E}\left[R_M\right] - R_f\right). \tag{11.5}$$

The term I_G denotes the information available for the fund manager. At market equilibrium, for the CAPM, A_P must be null. It will be so if the fund manager has private information and good anticipations. Note that the new alphas of portfolios $S - C$ and $S + P$ are null for all strikes.

11.2.2.2.2 Strategy which maximizes the Sharpe ratio Recall that the Sharpe ratio for the market index is equal to 0.4743. The following figure presents the Sharpe ratio of portfolios which contain one unit of index and $-a$ units of call, as function of the strike K using the same financial parameter values as before.

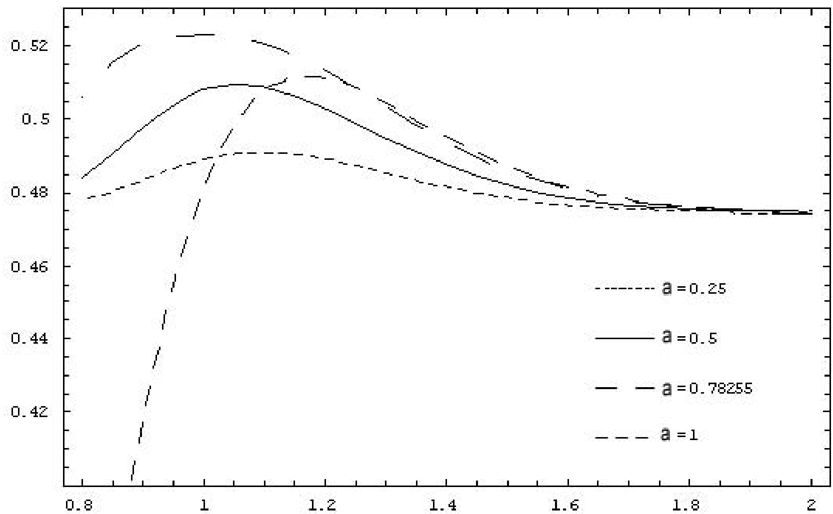


FIGURE 11.4: Sharpe ratio as a function of the strike K

The maximal Sharpe ratio is reached for the following values:

$$a = 0.7826 \text{ and } K = 0.992.$$

This is equal to 0.5251 and so is higher than that of the index.

Instead of searching for performance measures which cannot be manipulated by strategies based on options, Goetzmann *et al.* [257] propose to determine strategies which maximize the Sharpe ratio without requiring any skill.

They prove that the best static strategy has a probability distribution which is right truncated and has a left fat tail. Such a strategy can be approximated by a combination of a put and a call.

The portfolio is based on an investment of one unit in the index, on the selling of a European calls with strikes K, and on the purchase of b European puts with strikes H and $(K > H)$.

The portfolio value at maturity is given by:

$$V_T = S_T - a.(S_T - K)^+ + b.(H - S_T)^+. \tag{11.6}$$

Thus, its initial value is equal to:

$$P_0 = 1 - a.C\,(1, T, \sigma, r; K) + b.P\,(1, T, \sigma, r; H)\,. \tag{11.7}$$

At maturity, the profile is concave as shown by figure 11.5, which corresponds to portfolios maximizing the Sharpe ratio.

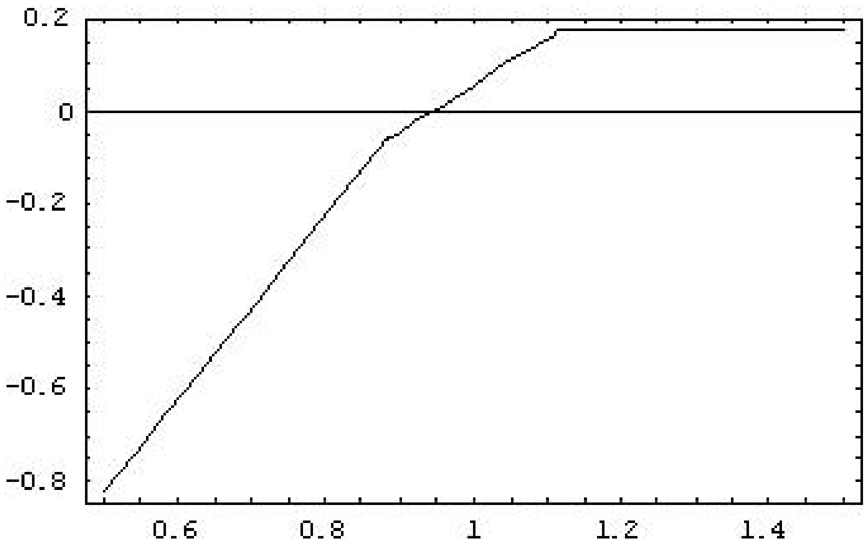


FIGURE 11.5: Portfolio profile maximizing the Sharpe ratio

Using now a units of call and b units of put, we get a maximal Sharpe ratio equal to 0.531. This new maximum corresponds to the following values:

$$a = 0.705,\ K = 0.846,\ \text{and}\ b = 1.93,\ H = 1.126.$$

Note that the main part of the Sharpe ratio increase is obtained from only one option.

These results can be compared with those shown in Chapter 10 when maximizing expected utility with insurance constraints.

In addition, Goetzmann *et al.* [257] show that the Sharpe ratio is not very sensitive to the choice of strikes. Therefore, options close to the spot index price can be used since they are more liquid.

To summarize, if we are only interested by the performance measure, results of Goetzmann *et al.*[257] suggest that the probability distribution of a fund which has a high Sharpe ratio, would be compared with the strategy which maximizes the Sharpe ratio. In particular, this is true when we consider the performance measure of hedge funds based on style analysis.

11.2.3 Alternative performance measures

Chapter 2 deals with the choice of appropriate risk measures. In particular, downside risk measures are analyzed. Performance measures for hedge funds can be based on such risk measures which focus on loss risk.

11.2.3.1 The semi-variance

The Sharpe ratio is based on a dispersion risk measure around the mean. Therefore, it does not allow us to determine if the variations are below or above the mean. The semi-variance, SV, is a possible tool to avoid such a problem:

$$SV(R) = \mathbb{E}\left[\left[(\mathbb{E}[R] - R)^+\right]^2\right]. \tag{11.8}$$

The semi-variance takes account of such assymetry. Lower partial moments are extensions of the semi-variance. They are defined by:

$$\mathbb{E}[\left[(\mathbb{E}[R] - R)^+\right]^p], \tag{11.9}$$

where the values of parameter p are above 2, which allows us to take better account of assymetries and potential high risks.

11.2.3.2 Sortino ratio

This is one of the most famous ratios. It takes account of loss expectations ("downside risk"). It is defined by (see Sortino et al. [475], [476], [477]):

$$Sor(R) = \frac{\mathbb{E}[R] - L}{\sqrt{\mathbb{E}[[(L-R)^+]^2]}}, \tag{11.10}$$

where L denotes the minimal acceptable return level (MAR).

This ratio is based on the same principle as the Sharpe ratio. However, the riskless rate is replaced by the level L, that is by the minimal acceptable return level, while the standard deviation of the return $\sigma(R)$ is replaced by the standard deviation of those returns that are below the level L.

The choice of the level L can be made according to different criteria:

- In order to control the loss risk, we can take $L = 0$.
- If we consider the riskless rate R_f as a benchmark, we can consider the value $L = R_f$.
- If we want to compare the performance of funds with each other, the level L can be chosen equal to the mean of return expectations of these funds.

As seen in what follows, (unfortunately) this choice is crucial to rank the funds.

11.2.3.3 The Omega performance measure

This performance measure was introduced by Keating and Shadwick [326]. Contrary to standard performance measures, such as the Treynor and Sharpe ratios or the Jensen alpha, it takes account of the whole probability distribution.

The Omega measure considers both the gain and loss probabilities. It is defined by:

$$\Omega_F(L) = \frac{\int_L^b (1 - F(x))\, dx}{\int_a^L F(x)\, dx}. \tag{11.11}$$

The function $F(.)$ is the cumulative distribution function of the financial assets with range (a, b) and w.r.t. the probability distribution $\mathbb{P}$ and the reference level L chosen by the investor.

For a given level, the investor would always prefer the portfolio with the highest Omega value.

The Omega ratio can be written as:

$$\Omega_{F_X}(L) = \frac{\mathbb{E}_{\mathbb{P}}\left[(X - L)^+\right]}{\mathbb{E}_{\mathbb{P}}\left[(L - X)^+\right]}. \tag{11.12}$$

REMARK 11.1 Therefore, the Omega function $\Omega_{F_X}(L)$ is the ratio of the expectation of gains above the level L on the expectation of losses below the level L. As mentioned in Kazemi *et al.* [325], the Omega ratio can be viewed as the ratio of a call on a put having the same strike L written on the same underlying asset (the portfolio value or its return) but with values computed w.r.t. the historical probability $\mathbb{P}$ instead of a risk-neutral one. $\square$

The Omega function also satisfies the following properties:

- For $L = \mathbb{E}_{\mathbb{P}}[X]$, $\Omega_{F_X}(L) = 1$.
- $\Omega_{F_X}(.)$ is a decreasing function.

  Thus, from the previous property, we deduce that:

  - For levels of L smaller than the mean, the Omega ratio is positive.

  - For values of X higher than the mean, the Omega ratio is negative.

- $\Omega_{F_X}(.) = \Omega_{G_X}(.)$ if and only if $F = G$.

  When portfolio returns always have identical Omega ratios, their probability distributions are equal.
- The Omega ratio is compatible with the second-order stochastic dominance:
$$X \succeq_2 Y \Longrightarrow \Omega_{F_X}(L) \geq \Omega_{F_Y}(L). \tag{11.13}$$
  As seen in Chapter 1, this is an interesting property.
- Kazemi *et al.* [325] define the Sharpe Omega ratio as follows:
$$Sharpe_{\Omega}(L) = \frac{\mathbb{E}_{\mathbb{P}}[X] - L}{\mathbb{E}_{\mathbb{P}}\left[(L-X)^{+}\right]} = \Omega_{F_X}(L) - 1. \tag{11.14}$$
  The upper term is the same as for the Sharpe ratio when the level L is equal to the riskless return. The risk measure $\mathbb{E}_{\mathbb{P}}\left[(L-X)^{+}\right]$ replaces the usual standard deviation.

Example 11.1
Consider the following standard case: The payoff X is the value at maturity T of a stock S which is assumed to follow a geometric Brownian motion:

$$X = S_0 \exp[(\mu - \sigma^2/2)T + \sigma W_T],$$

where $(W_t)_t$ is a standard Brownian motion which consequently is such that W_T has a Gaussian distribution $\mathcal{N}\left(0, \sqrt{T}\right)$.

Then, we have:
$$\mathbb{E}_{\mathbb{P}}[X] = S_0 \exp[\mu T].$$
Therefore, the mean return does not depend on the volatility.

Consequently:

- If $S_0 \exp[\mu T] < L$, then the Sharpe Omega ratio is an increasing function of the volatility (due to the vega of the put).

- If $S_0 \exp[\mu T] > L$, then the Sharpe Omega ratio is a decreasing function of the volatility.

Generally, the level L is chosen smaller than the mean (since it represents a loss w.r.t the expected value). Therefore, for the standard "buy-and-hold" strategy, the performance measure Omega is indeed decreasing w.r.t. the usual volatility risk.

$\square$

In the following numerical example, the Omega performance measure is applied to four hedge funds. The time period is 1992-2004.

Two types of hedge funds are examined:

1) *Hedge/Convertible Arbitrage:* A and D.

2) *Hedge/Equity Market Neutral:* B and C.

The estimation of parameters is made on monthly data.

TABLE 11.4: The four hedge funds characteristics

Characteristics	Fund A	Fund B	Fund C	Fund D
Mean	0.70	0.92	0.35	0.66
Median	0.80	0.80	0.27	0.58
Variance	0.54	0.51	11.74	2.13
Skewness	-0.74	0.65	0.35	0.56
Kurtosis	4.33	3.11	4.56	4.50

Their monthly returns are represented as follows:

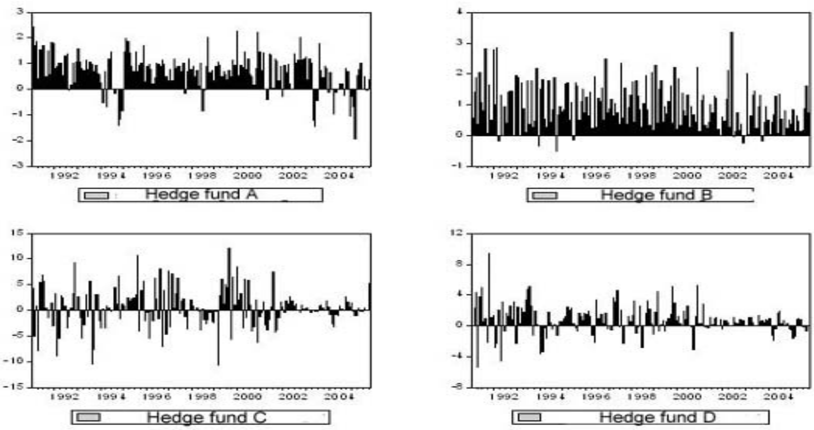


FIGURE 11.6: The monthly returns of the four hedge funds

The figure plots the Omega ratio as function of the reference level L:

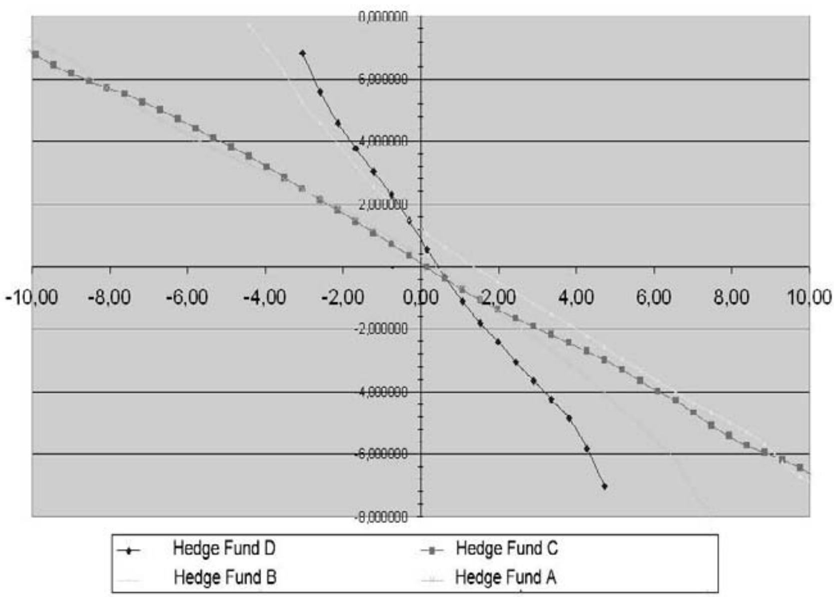


FIGURE 11.7: Omega ratio as function of the threshold level L

For this particular example, we can see that:

- None of the four funds dominate another one for all values of the threshold L (fund D for example dominates the three other ones for small values of L but it is dominated by the other ones for high values of L). However, fund B has a Omega ratio which is always above that of fund A, which may lead to stochastic dominance.

- Note that the choice of the reference level L is crucial, in particular around "rational" values such as the riskless rate where the ranking changes very quickly. Therefore, if the level of L is not an absolute reference, it seems necessary to define an additional criterion to determine this level. For this purpose, we can take into account the investor's risk aversion, the remuneration of the fund manager, or we can introduce weighted Omega ratios (judgment of financial experts, expectation for a given probability distribution on the level, *etc.*).

11.2.3.4 The ASRAP measure ("Alternative Style Risk Adjusted Performance")

In order to take into account the management style, Lobosco [360] proposes the "style risk adjusted performance" measure (SRAP).

The risk is measured by the volatility. Since hedge funds generally have asymetrical probability distributions with fat tails, it is necessary to examine at least their moments of orders 3 and 4.

For this purpose, the VaR developed by Cornish and Fisher [131] can be introduced to measure the risk. This is an extension of the standard VaR, which includes the *skewness* and the *excess kurtosis* of the return distribution.

The first step consists of calculating the VaR based on a Gaussian distribution then considering the extended VaR of Cornish-Fisher.

Define q by :

$$q = q_c + \frac{1}{6}(q_c^2 - 1)s + \frac{1}{24}(q_c^3 - 3q_c)k - \frac{1}{36}(2q_c^3 - 5q_c)s^2, \tag{11.15}$$

where q_c is the critical value at the probability level $(1-\alpha)$, s is the skewness, k is the excess kurtosis.

Then the adjusted VaR is given by:

$$VaR_{CF} = -(\mu - q\sigma), \tag{11.16}$$

where μ and σ are respectively the mean and the standard deviation of the probability distribution. Note that, if this distribution is Gaussian, then $s = 0$ and $K = 0$, so $q_c = q$. In that case, it is equal to the usual VaR.

We deduce that the ARAP measure is defined by:

$$\text{ARAP} = \frac{VaR_{CF}(I_{FOF})}{VaR_{CF}(HF)}\,(R_{HF} - R_f) + R_f, \tag{11.17}$$

where I_{FOF} is an index FOF, and HF denotes the hedge fund to be examined.

The SRAP measure is determined by the difference between the measure RAP of the portfolio and the measure RAP of the style benchmark, which corresponds to the portfolio style:

$$\text{ASRAP} = \text{ARAP(fund)} - \text{ARAP(style index)}. \tag{11.18}$$

11.2.4 Benchmarks for alternative investment

The alternative investment searches for absolute performance. However, the use of the riskless return as a benchmark is not always a convenient tool for all styles of hedge funds, in particular when their beta is not null, and if the implicit assumptions which validate the CAPM are not satisfied.

Nowadays, hedge fund performance is measured more and more relative to a style benchmark.

Such a benchmark must satisfy the following properties:

- Transparency - The list of hedge funds that are included in the benchmark must be detailed and the way to compute their performance must be specified.

- Representativeness - A large set of hedge funds must be considered while excluding those which are too small or the management of which is too hazardous.

- Weighting - The use of the respective capitalizations of different funds is not easy to handle. The recent development of many funds and their lack of standardization does not facilitate this weighting. Therefore, equal weights are often considered.

- Accessability of the hedge funds.

- The "reporting" frequency.

To satisfy some of the previous required properties, we can:

- Analyze the return of a hedge fund w.r.t. the return of a given portfolio having the same strategy, but in a passive way (*passive benchmark*). Agarwal and Naik ([9],[10]) introduce multifactorial models in order to analyze the types of assets and strategies used by the hedge funds.
  These factors include the strategies based on options on observable assets, as suggested in Agarwal and Naik [9]. This approach is further studied in Schneeweis and Spurgin [456]. They prove that the performance measure of hedge funds can actually be based on option strategies.

- Compare the return of a given fund with that of a representative index (*active benchmark*).

- Consider a pure style index. This type of index is defined as the true index which is not observable.

11.2.5 Measure of the performance persistence

Among studies about performance persistence of hedge funds, we have:

- Brown *et al.* [93] who examine if the future returns can be predicted from past observations. Their conclusion is that this is not the case.

- Agarwal and Naik ([7], [8]) conclude that there exists a short term performance. They examine annual returns during the time period 1982-1998. Their results prove that the performance persistence level is significant for a monthly observation, but it is reduced for annual observations.

Park and Staum [402] show that there exists a performance persistence for CTA's.

- Elton *et al.* [198] and Brown *et al.* [93] argue that the survival bias may induce performance persistence.

- Harri and Brorsen [282] show that some funds are actually performant over a sufficiently long time period.

- Baquero *et al.* [41] take account of the "look-ahead bias" which is induced by the multiperiodic sampling. This bias can induce a difference of about 3.8%. However they confirm the persistence.

The general conclusion is that there exists a performance persistence, for both the "losers" and the "winners." Note also that the performance is linked to management costs received by the fund managers, as noted by example Caglayan and Edwards [100].

Usual statistical methods are based on:

- Regression of the returns; in particular, the behavior of the alpha coefficient is examined.

- Style analysis.

- Correlation tests, such as Spearman's test.

11.3 Optimal allocation in hedge funds

The management of a fund which itself uses other funds such as hedge funds allows for diversification. This is due to the weak correlation of hedge funds with standard financial assets, as shown in the next figure, where a hedge fund index is compared with standard financial indexes.

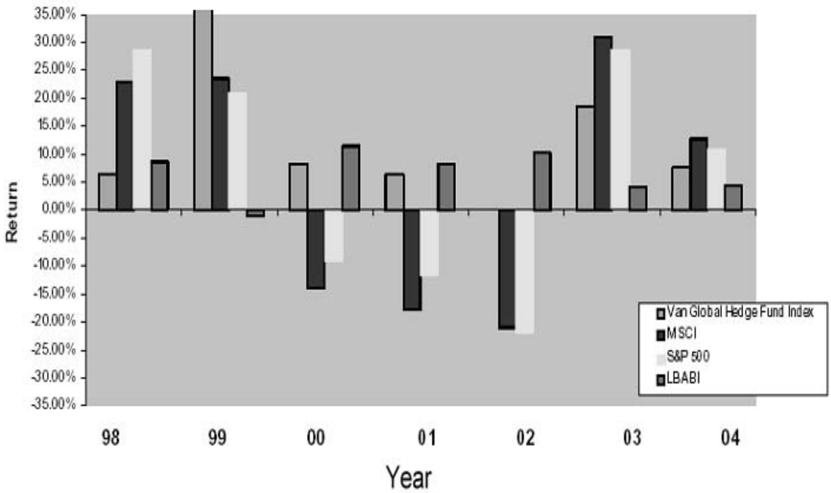


FIGURE 11.8: Correlation of hedge funds/standard funds

We can see that returns of hedge funds can evolve independently from traditional assets (confirmed by the correlation values).

What is the percentage of hedge funds to include in a fund or a portfolio? Obviously, the answer is not easy:

- Some funds cannot include more than a given percentage (for example, 10%). Otherwise, their category is changed.
- A minimal diversification is desirable to profit from weak *beta* and high *alpha*.

Note that nowadays most financial institutions include hedge funds. From a theoretical point of view, as a first step, a mean-variance analysis can be developed as in Schneeweis and Spurgin [456], who determine the mean-variance

frontier, including the S&P500, the fixed-income index Lehman Brothers, and a hedge fund index, the EACM 100. They conclude that hedge funds must be used. This analysis is further extended by Cvitanic et al. [141], who introduce a model based on a dynamic mean-variance criterion. In this framework, they prove that the model risk reduces the percentage invested in hedge funds. However, as mentioned by Lhabitant [355], the mean-variance criterion does not take account of moments with orders higher than 2, which strongly limits the analysis. It is the reason why for example Chabaane et al. [111] introduce various optimization criteria such as the expected shortfall.

11.4 Further reading

Lhabitant [354] provides an overview of hedge funds including description, classification, and performance. Fung and Hsieh [248] examine the hedge fund survival lifetimes, while in [249] they show how hedge fund strategies may involve risk. Brooks and Kat [91] examine statistical properties of hedge fund index returns and deduce their implications for portfolio managemen-t. The book of Gregoriou *et al.* [268] contains a survey about performance persistence. Bookstaber and Roger [83] highlight the problem of measuring performance of portfolios with options. Bacmann and Scholz [36] propose alternative performance measures for hedge funds. Bonnet and Nagot [81] propose a class of performance measures in order to evaluate alternative investment, regardless of assumptions on payoff. The representation of these measures involves the Log-Laplace transform of the asset distribution - among them: the squared Sharpe ratio, the Stutzer's rank ordering index and the Hodge's Generalized Sharpe ratio. Cascon *et al.* [107] provide some mathematical properties of the Omega measure.

Avouyi et al. [27] apply the Omega to the portfolio allocation choice problem, using Threshold Accepting. They introduce conditional copula to take into account non-Gaussian returns, extreme joint movements, time-varying dependence, and volatility. They apply these methods to a portfolio composed of three total stock market indices (US, UK and Germany). Amenc and Martellini [23] also examine portfolio optimization involving hedge funds using an improved estimator of the covariance structure of hedge fund index returns. Using data from CSFB-Tremont hedge fund indices, they conclude that ex-post volatility of minimum variance portfolios is between 1.5 and 6 times lower than that of a value-weighted benchmark. Thus, inclusion of hedge funds in a portfolio can potentially generate a dramatic decrease in the portfolio volatility without lower expected returns. Krokhmal *et al.* [337] also endeavor to optimize a portfolio of hedge funds. They examine linear rebalancing strategies using Value-at-Risk and CVaR criteria.

Appendix A

Appendix A: Arch Models

Linear and non linear processes

Time series have been introduced in particular to describe and to predict discrete-time dynamics. One of the most popular models is the so-called *autoregressive moving average process* (ARMA). The current value of the series is a linear function of its own lagged values and of current and past values of a "noise"process, usually called the innovation process. However, it is set-up in a linear framework (rather strong approximation), and no constraint is usually imposed on the moving average parameters (which does not allow taking into account structural relations). Financial time series, for example, exhibit non-linear dynamics and are submitted to structural constraints such as equilibrium conditions.

Therefore, a new family of time series have been introduced by Engle [204]: the ARCH (*Autoregressive Conditionally Heteroscedastic*) models. These models allow consideration of nonlinear time series models. They can also be applied to path dependent volatility models.

Weak and strong stationarity

A stationary time series is stationary if it has no trend and no seasonality. It is homogeneous with respect to time. More precisely:

DEFINITION A.1 *A time series $(X_t)_t$ is weakly stationary if:*

$$\mathbb{E}(X_1) = ... = \mathbb{E}(X_t) = ... = \mu,$$
$$Cov(X_t, X_{t+h}) = \gamma(h),\ \forall (t,h) \in \mathbb{N} \times \mathbb{Z}. \tag{A.1}$$

In particular, the autocovariances only depend on the time period h and not on current time t.

A time series $(X_t)_t$ is strongly stationary if for any n and any $t_1 < ... < t_n$, the probability distribution of the random vector $(X_{t_1}, ..., X_{t_n})$ depends only on n. Strong stationarity obviously implies weak stationarity.

Example A.1

A well-known non-stationary time series is the random walk (RW):

$$X_t = X_{t-1} + \varepsilon_t,$$

where $(\varepsilon_t)_t$ is a standard Gaussian white noise (WN) (*i.e.* the random variables ε_t are independent and have the same probability distribution, which is the standard Gaussian law $\mathcal{N}(0,1)$).

Such series "diverge." Consider for example a stationary time series which is a first-order autoregressive process (AR1):

$$Y_t = \phi . Y_{t-1} + \varepsilon_t, \text{ with } |\phi| < 1.$$

The behaviors of the two time series are illustrated in the next figure (for $X_0 = Y_0 = 0$).

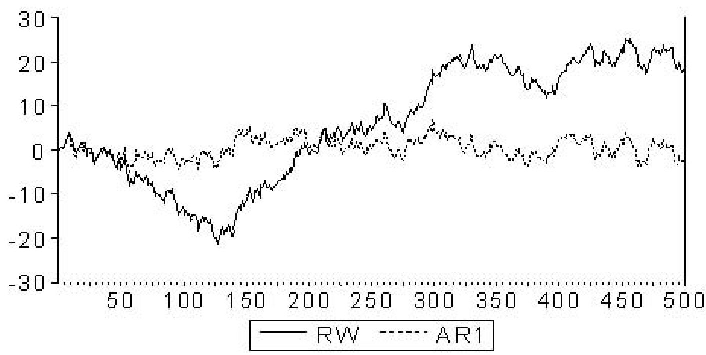


FIGURE A.1: Random walk (RW) and autoregressive process of order 1 ($AR1$)

The stationary series varies around its mean (equal to 0) whereas the random walk has a high variance since we have:

$$X_t = X_0 + \sum_{t=1}^{N} \varepsilon_t \Longrightarrow \sigma^2(X_t) = \sigma^2(\sum_{t=1}^{N} \varepsilon_t) = \sigma^2 N.$$

Note that the latter equality proves that the random walk is not stationary. □

Standard econometrics analysis is based on stationarity and is not adapted to non-stationary time series as shown by the following example of Granger and Newbold [264].

Example A.2 *Fallacious regressions*

Consider a linear regression between two independent random walks $RW1$ and $RW2$ such that (the numbers between braces are the Student's statistic values):

$$RW1_t = \underset{(7)}{3.5167} + \underset{(72.25)}{0.59}\, RW2_t + \varepsilon_t, \ \textit{with an } R^2 = 0.72.$$

We note that the variable $RW2$ would allow us to explain the dynamics of $RW1$: there exists a linear relation between these two random variables, whereas they are independent. This surprising result is due to their common propensity for diverging. However, if we consider two independent Gaussian white noises $\varepsilon_t^{(1)}$ and $\varepsilon_t^{(2)}$ (thus stationary processes), no linear relation between them appears:

$$\varepsilon_t^{(1)} = \underset{(-0.66)}{-0.0148} - \underset{(-0.841)}{0.086}\, \varepsilon_t^{(2)} + \varepsilon_t, \ \textit{with an } R^2 = 0.018.$$

None of the previous coefficients is significant. □

ARMA processes

The current value of X_t is a linear function of the past values of X and of past and current values of a white noise process ε. Define:

$$\underline{X}_{t-1} = (X_1, ..., X_{t-1}).$$

Denote also by $L\mathbb{E}\left[X_t \left|\underline{X}_{t-1}\right.\right]$ the linear regression of X_t (the best prediction of X_t by means of a linear function of $X_1, ..., X_{t-1}$).

DEFINITION A.2 *A second order stochastic process is:*
1) an autoregressive process of order K if and only if:

$$\mathbb{E}\left[X_t \left|\underline{X}_{t-1}\right.\right] = \mathbb{E}\left[X_t \left|X_{t-1}, .., X_{t-K}\right.\right], \forall t.$$

2) a linear autoregressive process of order K if and only if:

$$L\mathbb{E}\left[X_t \left|\underline{X}_{t-1}\right.\right] = L\mathbb{E}\left[X_t \left|X_{t-1}, .., X_{t-K}\right.\right], \forall t.$$

An ARMA representation is defined as follows:

$$X_t = C + \Phi_1 X_{t-1} + ... + \Phi_p X_{t-p} + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - ... - \Theta_q \varepsilon_{t-q}, \tag{A.2}$$

where $\Phi_1, ..., \Phi_p$ and $\Theta_1, ..., \Theta_q$ are square matrices and C is a vector.

The autoregressive and moving average lag-polynomials are given by:

$$\begin{aligned} \Phi(L) &= Id + \Phi_1 L - ... - \Phi_p \mathbb{L}^p, \\ \Theta(L) &= Id - \Theta_1 L - ... - \Theta_q L^q, \end{aligned} \tag{A.3}$$

where L denotes the lag operator, $L(X_t) = X_{t-1}$. The ARMA representation is then defined by:

$$\Phi(L)(X_t) = C + \Theta(L)\varepsilon_t. \tag{A.4}$$

The coefficients $\Phi_1, ..., \Phi_p$ and $\Theta_1, ..., \Theta_q$ are usually constrained. For example, the roots of the equations:

$$\det(\Phi(z)) = 0 \text{ and } \det(\Theta(z)) = 0$$

are supposed to be outside the unit circle ($|z| > 1$). Indeed, under this latter assumption, the polynomials $\det(\Phi(z))$ and $\det(\Theta(z))$ can be inverted. This leads to alternative representations of the process X:

$$\text{(infinite moving average representation) } X_t = \Phi(L)^{-1}C + \Phi(L)^{-1}\Theta(L)\varepsilon_t,$$
$$\text{(infinite autoregressive representation) } \Theta(L)^{-1}\Phi(L)X_t = \Theta(1)^{-1}C + \varepsilon_t.$$

Thus ARMA processes are relatively tractable. Note that they can be considered as truncated approximations of weakly stationary processes due to Wold's decomposition.

THEOREM A.1 **Wold's theorem**

Consider a weakly stationary process $(X_t)_t$ such that

$$\lim_{k\to\infty} \mathbb{E}\left[X_{t+k} \,|X_t\right] = \mathbb{E}\left[X_t\right],$$

which means that no information at an infinite horizon is given from the observations prior to time t.

Then, this process always has an infinite moving average representation:

$$X_t = C_0 + A(L)\varepsilon_t, \tag{A.5}$$

where $(\varepsilon_t)_t$ is a sequence of homoscedastic noise variables, $V\varepsilon_t = \Omega$, which are uncorrelated, $Cov(\varepsilon_t, \varepsilon_{t'}) = 0$, $\forall t, \forall t'$, with zero mean, $\mathbb{E}\left[\varepsilon_t\right] = 0$, and such the coefficients $A_1, ...A_j, ...$satisfy the following stability condition:

$$\sum_{j=0}^{\infty} A_j \Omega A_j' < \infty.$$

Note that the process ε satisfies:

$$\varepsilon_t = X_t - L\mathbb{E}\left[X_t \,|\varepsilon_t\right],$$

which is the reason why this process is called the innovation process.

ARCH model

The Autoregressive Conditionally Heteroscedastic model takes account of time dependent conditional variances. For example:

$$X_t = \varepsilon_{t-1}^2 \varepsilon_t,$$

where ε is a Gaussian white noise with variance σ^2. The process X is weakly stationary and the conditional variance depends on lagged residuals:

$$Var\left(X_t \left| \underline{X}_{t-1}\right.\right) = Var\left(\varepsilon_{t-1}^2 \varepsilon_t \left| \underline{X}_{t-1}\right.\right) = \sigma^2 \varepsilon_{t-1}^2.$$

ARCH volatility models

The univariate case

A standard ARCH volatility process of order p can be written as:

$$R_t = \sum_{n=1}^{k} \Phi_n R_{t-n} + \epsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2, \tag{A.6}$$

where σ_t^2 is the conditional variance of the innovation process ϵ_t. The previous representation is based on a moving average of the squares of the innovations. The coefficients $\alpha_0 > 0,\ \alpha_i \geqslant 0,\ i = 1, \ldots, p$ are assumed to be positive. Under the assumption $\sum_{i=1}^{p} \alpha_i < 1$, the time series is stationary.

This representation has been extended by Bollerslev *et al.* [80] who introduce the GARCH(p, q) model for which the conditional variance satisfies the following equation:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \tag{A.7}$$

In that case, the variance is defined as the sum of an autoregressive term and of a moving average of the squares of the innovations. The conditional variance is stationary if the condition$\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i < 1$ is satisfied.

More generally, the term $\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i$ indicates the persistence degree of the variance, since the non-conditional variance can be written as:

$$\sigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i}.$$

When $\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i = 1$, the non-conditional variance is infinite. The conditional variance follows an IGARCH process (*Integrated GARCH*).

Note that if the standardized innovations ($z_t = \frac{\epsilon_t}{\sigma_t}$) have a standard Gaussian distribution, then the marginal distribution of the innovation process ϵ_t is characterized by a kurtosis which is always higher than the standard Gaussian one. In order to fit fat tail distributions, other innovation processes can be introduced, for example GED or Student distributions.

In order to take asymmetrical distributions into account, several models have been introduced by Nelson [398], Engle and Ng [208], Glosten *et al.* [256], and Zakoian [509], *etc.*

- GJR model (Glosten, Jagannathan, and Runkel [256]) defined by:

$$\sigma_t^2 = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \gamma I_{t-1}\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2,$$

  with $\alpha_0 > 0$, $\alpha \geqslant 0$, $\alpha + \gamma \geqslant 0$ and $\beta \geqslant 0$. The process is stationary if $\beta + \alpha + \gamma/2 < 1$.

- TGARCH model (*Threshold GARCH*), Zakoian [509] defined by:

$$\sigma_t^2 = \alpha_0 + \alpha\left|\varepsilon_{t-1}\right| + \gamma I_{t-1}\left|\varepsilon_{t-1}\right| + \beta\sigma_{t-1},$$

  with $\alpha_0 > 0$, $\alpha \geqslant 0$, $\alpha + \gamma \geqslant 0$ and $\beta \geqslant 0$. The volatility process is stationary if:

$$\beta^2 + \frac{\alpha^2 + (\alpha+\gamma)^2}{2} + 2\beta\frac{2\alpha+\gamma}{\sqrt{2\pi}} < 1.$$

- EGARCH model (*Exponential GARCH*) (Nelson [398]) defined by:

$$\ln{\sigma_t}^2 = \alpha_0 + \alpha(\left|z_{t-1}\right| - E\left|z_{t-1}\right|) + \gamma z_{t-1} + \beta\ln\sigma_{t-1}^2,$$

  where $E\left|z_{t-1}\right| = \sqrt{2/\pi}$ under the normality assumption.

  The volatility process is stationary if $\beta < 1$.

- If the long term volatility is non constant, we can consider "component-models."

  These models allow us to take into account mean-reverting properties:

$$\sigma_t^2 - q_t = \alpha\left(\varepsilon_{t-1}^2 - q_{t-1}\right) + \beta\left(\sigma_{t-1}^2 - q_{t-1}\right), \tag{A.8}$$
$$q_t = \omega + \rho\left(q_{t-1} - \omega\right) + \phi\left(\varepsilon_{t-1}^2 - \sigma_{t-1}^2\right). \tag{A.9}$$

  The term q_t is the long term volatility component. Equation (A.8) describes the transient volatility component, $\sigma_t^2 - q_t$, which converges to zero with speed $(\alpha + \beta)$.

  Equation (A.9) defines the long term volatility component q_t, which converges to ω with speed ρ.

- To take account of asymmetry, the TARCH model can be used:

$$\sigma_t^2 - q_t = \alpha\left(\varepsilon_{t-1}^2 - q_{t-1}\right) + \gamma\left(\varepsilon_{t-1}^2 - q_{t-1}\right) I_{t-1} + \beta\left(\sigma_{t-1}^2 - q_{t-1}\right). \tag{A.10}$$

In all previous relations, β is the autoregressive term, α is the effect of a shock on return, and γ is the asymmetry effect corresponding to an additional impact of a negative shock. Thus, for the GJR and TGARCH models, the effect of a positive shock is measured by α, and the effect of a negative shock is measured by $\alpha + \gamma$ (γ is assumed to be positive).

For the EGARCH model, the effect of a positive shock is measured by $\alpha + \gamma$ and the effect of a negative shock is measured by $\alpha - \gamma$. In that case, γ must be negative so that a negative shock has no higher impact than a positive shock.

REMARK A.1 The comparison of these asymmetrical models can be based on different response curves to innovations which illustrate the innovation effects on the conditional variance (*News Impact Curves*) proposed by Engle and Ng [208]. □

The multidimensional case

Consider a N-multidimensional GARCH model:

$$Y_t = \mu + \varepsilon_t, \text{ with } \varepsilon_t \mid I_{t-1} \rightsquigarrow N\left(0, H_t\right).$$

Two basic examples of such models are:

- Diagonal VECH:

$$H_t = A_0 + \sum_{i=1}^{p} A_i \otimes H_{t-i} + \sum_{i=1}^{q} B_i \otimes \varepsilon_{t-1}\varepsilon_{t-1}^T.$$

- BEKK (see [206]) :

$$H_t = A_0 A_0^T + \sum_{i=1}^{p} A_i H_{t-i} A_i^T + \sum_{i=1}^{q} B_i \left(\varepsilon_{t-1}\varepsilon_{t-1}^T\right) B_i^T.$$

Consider for example the multidimensional GARCH model which is called the DCC-MVGARCH (*Dynamic Conditional Correlations Multivariate Generalized Auto Regressive Conditional Heteroscedastic model*) introduced by Engle and Sheppard [209].

$$\begin{aligned} \epsilon_t \ \mid \ I_{t-1} &\rightsquigarrow N\left(0, H_t\right), \\ H_t &\equiv D_t R_t D_t, \end{aligned}$$

where ϵ_t are the residuals from the filtration, D_t is a diagonal matrix with coefficients which are stochastic standard deviations generated by univariate GARCH processes, and R_t denotes a stochastic correlation matrix.

The log-likelihood is defined by:

$$\begin{aligned}
L &= -\frac{1}{2}\sum_{t=1}^{T}\left(k\log(2\pi)+\log(|H_t|)+\epsilon_t' H_t^{-1}\epsilon_t\right),\\
&= -\frac{1}{2}\sum_{t=1}^{T}\left(k\log(2\pi)+\log(|D_t R_t D_t|)+\epsilon_t' D_t^{-1}R_t^{-1}D_t^{-1}\epsilon_t\right),\\
&= -\frac{1}{2}\sum_{t=1}^{T}\left(k\log(2\pi)+\log(|D_t|)+\log(|R_t|)+\eta_t' R_t^{-1}\eta_t\right),
\end{aligned}$$

where $\eta_t \rightsquigarrow N(0,R_t)$ are the residuals, standardized by their conditional standard deviations. In Engle and Sheppard [209], the coefficients of D_t are univariate GARCH processes given by:

$$H_{it} = \omega_i + \sum_{p=1}^{P_i}\alpha_{ip}\epsilon_{it-p}^2 + \sum_{p=1}^{Q_i}\beta_{iq}h_{it-p},$$

where H_{it} is the usual conditional GARCH variance and, for $i = 1, 2, ..., k$, usual non-negativity constraints hold together with the stationary condition:

$$\sum_{p=1}^{P_i}\alpha_{ip} + \sum_{q=1}^{Q_i}\beta_{iq}h_{it-p} < 1.$$

The dynamic correlations are:

$$\begin{aligned}
Q_t &= \left(1-\sum_{m=1}^{M}\alpha_m-\sum_{n=1}^{N}\beta_n\right)\overline{Q}+\sum_{m=1}^{M}\alpha_m\left(\eta_t-m\eta_{t-m}'\right)+\sum_{n=1}^{N}\beta_n Q_{t-n}\\
R_t &= Q_t^{*-1}Q_t Q_t^{*-1},
\end{aligned}$$

where α and β are the weights, $\overline{Q}$ is the non conditional covariance of standardized residuals, Q_t^* is a diagonal matrix with coefficients which are squares of the diagonal coefficients q_{ii} of the matrix Q_t, and M and N are the DCC lags. Note that the coefficients of R_t are given by:

$$\rho_{ijt} = \frac{q_{ijt}}{\sqrt{q_{ii}q_{jj}}}.$$

Appendix B

Appendix B: Stochastic Processes

Stochastic basis, filtration, stopping times

The notion of filtration is used to represent, for example, the flow of possible observations of prices on a financial market which is available for traders and portfolio managers.

DEFINITION B.1 *Suppose* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space.*
A filtration $(\mathcal{F}_t)_t$ *is an increasing family of sub-sigma-fields* $\mathcal{F}_t \subset \mathcal{F}$. *A* stochastic basis *is a probability space equipped with a filtration. Increasing means that if* $s \leq t$, *then* $\mathcal{F}_s \subset \mathcal{F}_t$. $\mathcal{F}_t$ *is usually interpreted as the set of events that occur before or at time* t. *Generally,* $\mathcal{F}_t$ *represents the history of some process observed up to time* t, *but other possible histories are allowed.*

It is assumed that the "usual conditions" are satisfied:

i) *Complete* : every $\mathbb{P}$-null set in $\mathcal{F}$ belongs to $\mathcal{F}_0$ and so to all $\mathcal{F}_t$.

ii) *Right continuous* : $\mathcal{F}_t = \cap_{s>t} \mathcal{F}_s$.

DEFINITION B.2 *A continuous time stochastic process* X *is a family of random variables* $(X_t)_t$ *defined on* $(\Omega, \mathcal{F}, \mathbb{P})$, *indexed by* t, *which take values in* $(E, \mathcal{E})$. *Hence, for all* t, X_t *is a random variable with values in* E. *Moreover, for each fixed* ω *(which represents a "state of the world"),* $t \to X_t(\omega)$ *is a function defined on* $[0, T]$, *called a* path *or a* trajectory *of the process* X.

The concept of *Martingale* is crucial in the modern theory of finance. If the price process M of an asset is a martingale, the conditional expectation at time s of the future value M_t of the stock at time t is given by its current value M_s.

DEFINITION B.3 *A* martingale *(resp.* submartingale*, resp.* supermartingale*) is an adapted process* X *on the basis* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ *whose paths*

are all right continuous and left limited (rcll) $\mathbb{P}-$almost surely, such that every X_t is integrable and such that for $s \leq t$:

$$X_s = \mathbb{E}_{\mathbb{P}}[X_t|\mathcal{F}_s] \text{ (resp. } X_s \leq \mathbb{E}_{\mathbb{P}}[X_t|\mathcal{F}_s], \text{resp. } X_s \geq \mathbb{E}_{\mathbb{P}}[X_t|\mathcal{F}_s]).$$

In order to model dynamics of financial assets, several types of stochastic processes are introduced: Brownian motion, Poisson processes, and, more generally, on Lévy processes, diffusions, diffusions with jumps, point processes, and martingales (see Shiryaev [469]).

Semimartingales and stochastic integrals

The class of semimartingales is the class of stochastic processes that is "rich" enough and sufficiently "tractable." It contains the previous processes and is stable under many of the usual transformations: localization, change of measure, change of filtration, and change of time. Finally, it is possible to define stochastic integrals with respect to semimartingales, which leads to the famous Stochastic Calculus and its very powerful results, like the Ito formula.

DEFINITION B.4 *1) A semimartingale is a process X which has the following decomposition (not unique)*

$$X = X_0 + A + M,$$

where X_0 is finite-valued and $\mathcal{F}_0-$measurable, M is a local martingale with $M_0 = 0$, and A has finite-variation.
2) A special semimartingale is a semimartingale X for which A is moreover predictable. Furthermore, this decomposition is unique and called the canonical decomposition of the special semimartingale X.

There exist several ways to construct stochastic integrals (see *e.g.*, Protter [419] for details).

If X has finite-variation and if H is a bounded process, the integral

$$H.X_t = \int_0^t H_s dX_s$$

is directly defined as the Stieljes integration path-by-path. But this construction excludes such fundamental processes as, for instance, the Brownian motion whose paths almost surely have no finite variations over each finite interval. Martingales in general, and also Markov processes, are similarly

excluded. This is a standard problem with semimartingales X having no finite-variation, since the measure $dX_s(\omega)$ is not defined.
So a special construction of a stochastic integral had to be developed while taking into account that stochastic integrals must be defined from limit of sums as usual integrals. For this purpose, restrictions must be made on both integrand and integrator.

DEFINITION B.5 *A process H is said to be simple predictable if H has a representation*

$$H_t = H_0 1_{\{0\}}(t) + \sum_{i=1}^{n} H_i 1_{(T_i, T_{i+1}]}(t), \tag{B.1}$$

where $0 = T_0 < T_1 < ... < T_{n+1} < \infty$ is a finite sequence of stopping times, $H_i \in \mathcal{F}_{T_i}$ with $|H_i| < \infty$ a.s., and $0 \leq i \leq n$.

DEFINITION B.6 *Let X be a right continuous and left limited process. Define the linear mapping associated to stochastic integrals with respect to process X by, for any previous process H :*

$$J_X(H) = H_0 X_0 + \sum_{i=1}^{n} H_i (X^{T_{i+1}} - X^{T_i}), \tag{B.2}$$

whenever

$$H_t = H_0 1_{\{0\}}(t) + \sum_{i=1}^{n} H_i 1_{(T_i, T_{i+1}]}(t),$$

where $0 = T_0 < T_1 < ... < T_{n+1} < \infty$ is a finite sequence of stopping times, $H_i \in \mathcal{F}_{T_i}$, with $|H_i| < \infty$ a.s., and $0 \leq i \leq n$.

$J_X(H)$ is called the stochastic integral of H with respect to X.

Example B.1
Consider the standard Brownian motion W. Recall that W can be characterized as a continuous stochastic process with independent and (Gaussian) stationary increments (i.e. for each $s < t$, $W_t - W_s$ is independent of $\mathcal{F}_s$ and its distribution is the Gaussian law with mean 0 and variance equal to t) with $W_0 = 0$.
Let $(\mathbb{P}i_n)_n$ be a refining sequence (i.e. $\mathbb{P}i_n \subset \mathbb{P}i_m$ if $m > n$) of partitions of $[0, \infty)$ with mesh sizes converging to 0 as n goes to infinity. Consider

$$W_t^n = \sum_{t_k \in \mathbb{P}i_n} 1 1_{(t_k, t_{k+1}]}.$$

W_n converges to W. Now fix t and assume that t is a partition point of each $\mathbb{P}i_n$. Then,

$$J_W(W_n) = \sum_{t_k \in \mathbb{P}i_n, t_k < t} W_{t_k}(W^{t_{k+1}} - W^{t_k}),$$

and

$$J_W(W)_t = \lim_{n\to\infty} J_W(W_n)_t = \sum_{t_k \in \mathbb{P}i_n, t_k < t} W_{t_k}(W_{t_{k+1}} - W_{t_k}),$$

$$= \lim_{n\to\infty} \left\{ \sum_{t_k \in \mathbb{P}i_n, t_k < t} 1/2(W_{t_k} + W_{t_{k+1}})(W_{t_{k+1}} - W_{t_k}) - 1/2(W_{t_{k+1}} - W_{t_k})^2 \right\}$$

$$= 1/2\, W_t^2 - 1/2 \lim_{n\to\infty} \left\{ \sum_{t_k \in \mathbb{P}i_n, t_k < t} (W_{t_{k+1}} - W_{t_k})^2 \right\}.$$

Now, examine the last term on the right. Denote

$$Y_k = (W_{t_{k+1}} - W_{t_k})^2 - (t_{k+1} - t_k).$$

The sequence $(Y_k)_k$ is an iid sequence of random variables with zero mean. Thus

$$\mathbb{E}\left[\sum_{t_k \in \mathbb{P}i_n} (W_{t_{k+1}} - W_{t_k})^2 - (t_{k+1} - t_k) \right] = \sum_k \mathbb{E}[Y_k^2].$$

Moreover, $\left(\mathcal{F}rac W_{t_{k+1}} - W_{t_k} t_{k+1} - t_k\right)^2$ has the distribution of the square of a Gaussian random variable Z with zero mean and variance equal to 1. Therefore

$$\mathbb{E}[\sum_{t_k \in \mathbb{P}i_n} (W_{t_{k+1}} - W_{t_k})^2 - (t_{k+1} - t_k)] \le \mathbb{E}[(Z^2 - 1)^2]\, \mathrm{mesh}(\mathbb{P}i_n)t,$$

which converges to 0 as n goes to infinity. Thus, the last term on the right converges to t ($\mathbb{L}^2$ convergence, but almost-surely convergence can also be proved (see Protter [419])). Consequently,

$$\int_0^t W_s dW_s = 1/2\, W_t^2 - 1/2\, t,$$

which distinctly differs from the Riemann-Stieljes integral formula.

For processes A with continuous paths of finite variations, the term $\sum_{t_k \in \mathbb{P}i_n} (A_{t_{k+1}} - A_{t_k})^2$ converges to 0, which explains the difference between path-by-path Riemann-Stieljes integrals with respect to processes A and stochastic integral with respect to processes such as the Brownian motion. ▯

The quadratic variation of a semimartingale is a very convenient tool.

DEFINITION B.7 *1) The quadratic co-variation of two semimartingales X and Y, denoted by $[X, Y]$, is defined by:*

$$[X, Y] = XY - X_0 Y_0 - X_-.Y - Y_-.X \,. \tag{B.3}$$

2) The quadratic variation of a semimartingale X is $[X, X]$ and so:

$$[X, X]_t = X_t^2 - X_0^2 - 2\int_0^t X_{s-} dX_s. \tag{B.4}$$

Ito's formula for semimartingales

In what follows, $D_i f$ denotes the first partial derivative of the function f w.r.t. x_i and $D_{ij} f$ denotes the second partial derivative of the function f w.r.t. x_i and x_j.

THEOREM B.1
Let $X = (X^1, ..., X^d)$ be a d-dimensional semimartingale and f a class C^2 twice continuously-differentiable function on $\mathbb{R}^d$. Then $f(X)$ is a semimartingale and:

$$f(X_t) = f(X_0) + \sum_{i \leq d} D_i f(X_-).X^i + 1/2 \sum_{i,j \leq d} D_{ij} f(X_-).\langle X^{i,c}, X^{j,c} \rangle$$

$$+ \sum_{s \leq t} \left[f(X_s) - f(X_{s-}) - \sum_{i \leq d} D_i f(X_{s-}) \Delta X_s^i \right].$$

Doléans-Dade exponential formula

This notion is very useful to the financial theory, since most asset price processes are in fact Doléans-Dade exponentials. Moreover, it is also relevant to change of measures to compute, for example, the Radon-Nikodym derivatives of the risk-neutral probabilities.

Consider the equation

$$Y = 1 + Y_-.X \text{ (or equivalently } dY = Y_- dX \text{ and } Y_0 = 1),$$

where X is a given semimartingale and Y an unknown rcll adapted process.

By analogy with the ordinary differential equation $\frac{dy}{dx} = y$, Y is called the exponential of X.

THEOREM B.2

If X is a semimartingale then the above equation has one and only one rcll adapted solution (up to indistinguishability) which is a semimartingale, denoted by $\mathcal{E}(X)$, and given by

$$\mathcal{E}(X)_t = exp\left(X_t - X_0 - 1/2\langle X^c, X^c\rangle\right) \prod_{s\leq t}(1+\Delta X_s)e^{-\Delta X_s}\,. \tag{B.5}$$

Recall some basic properties of this exponential:

PROPOSITION B.1

1) If X has finite variation, then so has $\mathcal{E}(X)$ which is given by

$$\mathcal{E}(X)_t = exp\left(X_t - X_0\right) \prod_{s\leq t}(1+\Delta X_s)e^{-\Delta X_s}\,.$$

2) If X is a local martingale then so is $\mathcal{E}(X)$.
3) Let $T = inf\{t : \Delta X_t = -1\}$. Then, $\mathcal{E}(X) \neq 0$ on $[[0,T[[$, $\mathcal{E}(X_-) \neq 0$ on $[[0,T]]$ and $\mathcal{E}(X) = 0$ on $[[T,\infty[[$.

Note also the following useful property:

PROPOSITION B.2

Let X and Y be two semimartingales with $X_0 = Y_0 = 0$. Then

$$\mathcal{E}(X)\mathcal{E}(Y) = \mathcal{E}(X + Y + [X,Y])\,. \tag{B.6}$$

Example B.2

As in the Black and Scholes model, consider a rate of return X described by a Brownian motion with drift. Thus, we have :

$$X_t = \mu t + \sigma W_t,$$

where μ and σ are constants, and $(W_t)_t$ is a standard Brownian motion. Then, the stock price process S is solution of the equation $S = S_0 + S_-.X$. Thus S is the Doléans-Dade exponential of X and, since $\langle \sigma W, \sigma W\rangle_t = \sigma^2 t$, one obtains the very well-known geometric Brownian motion

$$S_t = S_0 \exp\left((\mu - 1/2\sigma^2)t + \sigma W_t\right).$$

□

Markov processes and stochastic differential equations

This paragraph is devoted to the Markov property of solutions to stochastic differential equations. It contains a brief survey on infinitesimal generators for diffusions.

One of the motivations for the development of stochastic integrals was the study of diffusions (i.e. the continuous strong Markov processes) as solutions of differential equations of the form:

$$X_t = X_0 + \int_0^t f(s, X_s)ds + \int_0^t g(s, X_s)dW_s,$$

where $(W_t)_t$ is a standard Brownian motion, and f and g are sufficiently regular functions to define a unique continuous solution which is strong Markov. From the general concept of semimartingale differentials, the terms ds and dW_s can be replaced by general semimartingales that should have independent increments so that the solution is Markovian. As mentioned in Protter [419], a "naive" definition of a Markovian process looks like a weakening of the property of independent increments:

DEFINITION B.8 *A process X with values in $\mathbb{R}^d$ and adapted is said to be a simple Markov process with respect to the filtration $(\mathcal{F}_t)_t$, if for each $s \geq 0$ the $\sigma-$fields $\mathcal{F}_s$ and $\sigma(X_t,\ t \geq s)$ (information uniquely from time s) are conditionally independent given X_s.*

This is in fact equivalent to: for $t \geq s$ and for every f bounded and Borel measurable,

$$\mathbb{E}[f(X_t)|\mathcal{F}_s] = \mathbb{E}[f(X_t)|\sigma(X_s)]. \tag{B.7}$$

"The best prediction of the future given the past and the present is the present."

From the previous relation, one can define a *transition function* for a Markov process as follows : for $s < t$ and f bounded and Borel measurable,

$$P_{s,t}(X_s, f) = \mathbb{E}[f(X_t)|\mathcal{F}_s]\,. \tag{B.8}$$

Letting $f(x) = 1\!\!1_C(x)$, the preceding relation reduces to

$$\mathbb{P}(X_t \in C|\mathcal{F}_s) = P_{s,t}(X_s, 1_C). \tag{B.9}$$

also denoted by $P_{s,t}(X_s, C)$.

If for any $s < t$, the transition function satisfies the relationship

$$P_{s,t} = P_{0,t-s}, \text{also denoted by } P(t-s)\,, \tag{B.10}$$

the Markov process is said to be time homogeneous and the transitions functions are a semigroup of operators, known as the transition semigroup$(P(t))_t$. In the time homogeneous case, the Markov property becomes, for any $u \geq 0$:

$$\mathbb{P}(X_{t+u} \in C | \mathcal{F}_t) = P(u, X_t, C). \tag{B.11}$$

Thus, a function $P(t, x, C)$ defined on $(0, \infty) \times \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d)$ is a time homogeneous transition function if:
1) For each (t, x), $P(t, x, .)$ is a probability on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.
2) For each x, $P(0, x, .)$ is the Dirac measure $\delta_x(C)$ at x (since $X_0 = x$).
3) For each $C \in \mathcal{B}(\mathbb{R}^d)$, $P(., ., C)$ is real-valued, Borel measurable, and bounded.
4) $P(t, x, C)$ satisfies the following relation, called the Chapman-Kolmogorov property: for each (t, x, C), t and $u \geq 0$:

$$P(t + u, x, C) = \int P(u, y, C) P(t, x, dy). \tag{B.12}$$

The probability measure m defined on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, by $m(A) = \mathbb{P}[X_0 \in C]$ is called the initial distribution of the process X.

A transition function for X and the initial distribution m determine the finite-dimensional distributions of X by

$$\begin{gathered}\mathbb{P}(X_0 \in C_0, X_{t_1} \in C_1 ..., X_{t_n} \in C_n) = \\ \int_{C_0} ... \int_{C_{n-1}} P(t_n - t_{n-1}, y_{n-1}, C_n) ... P(t_1, y_0, dy_1) m(dy_0) \,.\end{gathered} \tag{B.13}$$

It can be required that the Markov Property holds for stopping times.

DEFINITION B.9 *A time homogeneous simple Markov process X is Strong Markov if for any stopping time T with $\mathbb{P}[T < \infty] = 1$ and $u \geq 0$,*

$$\mathbb{P}(X_{T+u} \in C | \mathcal{F}_T) = P(u, X_T, C)\,, \tag{B.14}$$

or equivalently, for any function f bounded and Borel measurable

$$\mathbb{E}[f(X_{T+u}) | \mathcal{F}_T) = P(u, X_T, f)\,. \tag{B.15}$$

In the previous definition, the strong Markov property has been defined for only time homogeneous processes but, if X_t is a $\mathbb{R}^d$-valued simple Markov process, then $Y_t = (X_t, t)$ is a $\mathbb{R}^{d+1}$-valued simple Markov process.

Such examples of Markov processes are the Brownian motion and the Poisson process and typical solutions of stochastic differential equations.

A first standard example of such stochastic differential equations is given by the Ito diffusion processes. Consider the case of the evolution of a stock price subject to random shocks (resulting, for instance, from a multitude of

stock trade orders). If $b(t,x)$ is the trend of the price at the point x and time t (a kind of "velocity"), then to describe the value X_t at time t, it may be reasonable (under some assumptions like independence of the shocks) to use an equation like

$$\frac{dX_t}{dt} = b(t, X_t)dt + \sigma(t, X_t)\epsilon_t \ ,$$

where $(\epsilon_t)_t$ is a white noise and $\sigma(t,x)$ measures the amplitude of the shocks (the very well-known volatility in finance).

The mathematical interpretation of this equation, due to Ito is

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \ , \tag{B.16}$$

where $(W_t)_t$ is a Brownian motion.

This equation can be immediately extended to the case X is $\mathbb{R}^d$-valued and $(W_t)_t$ is a p-dimensional Brownian motion. $b(t,x) \in \mathbb{R}^d$ is usually called the drift, and $\sigma(t,x) \in \mathbb{R}^{d\times p}$ the diffusion coefficient.

In fact, since the works of Ito, the above equation is written with the use of stochastic integrals : if X_0 is given, then the solution is

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s \ , \tag{B.17}$$

where $(W_t)_t$ is a Brownian motion, and $b(t,x)$, $\sigma(t,x)$ are appropriately smooth to ensure the existence and the uniqueness of solutions.
Intuitively speaking, if $(\mathcal{F}_t)_t$ is the underlying filtration of the Brownian motion, then for small ϵ and for all $i, j \le n$,

$$\mathbb{E}[X^i_{t+\epsilon} - X^i_t | \mathcal{F}_t] = b^i(t, X_t)\epsilon + o(\epsilon) \ ,$$

$$\mathbb{E}[(X^i_{t+\epsilon} - X^i_t - b^i(t, X_t)\epsilon)(X^j_{t+\epsilon} - X^j_t - b^j(t, X_t)\epsilon)|\mathcal{F}_t] = \sigma^{'i}\sigma^j(t, X_t)\epsilon + o(\epsilon)$$

(notation: σ' denotes the transpose of the matrix σ)

The basic properties of Ito diffusions are :

1) The Markov property.
2) The strong Markov property.
3) The generator A of the process X can be expressed in terms of b and σ.

Two conditions are usually introduced on the coefficient functions b and σ to guarantee the existence and the uniqueness of the solution. For example,

THEOREM B.3
Let $T > 0$, $b(.,.) : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma(.,.) : [0,T] \times \mathbb{R}^{d\times p} \to \mathbb{R}^d$ be measurable functions satisfying

1)

$$|b(t,x)| + \| \sigma(t,x) \| \leq C(1+|x|)\ , \tag{B.18}$$

for some constant C (where $|.|$ is the Euclidian norm on $\mathbb{R}^d$, and $\| \sigma \|^2 = \sum \sigma_{ij}^2$) and the functions $t \rightarrow b(t,x)$ $t \rightarrow \sigma(t,x)$ are continuous.
2)

$$|b(t,y) - b(t,x)| + \| \sigma(t,y) - \sigma(t,x) \| \leq D|y-x|\ , \tag{B.19}$$

for some constant D.
Let Z be a random variable which is independent of the $\sigma-$field $\mathcal{F}^W$ generated by the Brownian W and such that $\mathbb{E}[Z^2] < \infty$. Then the stochastic differential equation

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t\ , t \leq T,\ with\ X_0 = Z$$

has a unique solution X which is continuous with respect to t and with components X^i that are adapted and satisfy $\mathbb{E}[Sup_{t \leq T}(X_t^i)^2] < \infty$.

The first assumption is called a linear growth condition. The second is a locally Lipschitz condition.

Two standard examples of diffusions are described in what follows.

Example B.3 **Diffusions**
1) The stochastic exponential $e^{W_t - 1/2t}$ is a diffusion with $b(t,x) = 0$ and $\sigma(t,x) = x$.

2) The Ornstein-Uhlenbeck process can be defined as follows:

$$dX_t = k(m - X_t)dt + \sigma dW_t,$$

where k is a non-negative constant and X_0 is independent of the Brownian W. Note that X is a Gaussian process (all finite-dimensional distributions are Gaussian). This process is often used in the term structure modeling since it has the mean-reverting property. In fact, since

$$X_t = m + (X_0 - m)e^{-kt} + \sigma \int_0^t e^{-k(t-s)} dW_t\ ,$$

then $\mathbb{E}[X_t] = m + (X_0 - m)e^{-kt}$ which tends to m as t goes to infinity.

$\square$

In fact, the definition of a diffusion is not standardized. For example, a process may be called a diffusion if it has continuous sample paths and if it satisfies the strong Markov property. Under the assumptions of the previous

theorem, the unique solution of the stochastic differential equation is a strong Markov process and, by construction, its paths are continuous.

Examine now the generator of a diffusion, solution of the stochastic differential equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \,, t \leq T, \text{ with } X_0 = Z \,.$$

Then, we deduce :

PROPOSITION B.3

If f is twice continuously differentiable functions with a compact support then f is in the domain $\mathcal{D}(A)$ of the operator A and

$$Af(x) = \frac{\partial f}{\partial t} + \sum_{i \leq d} b_i(x) \frac{\partial f}{\partial x_i} + 1/2 \sum_{i,j \leq d} (\sigma\sigma')_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j} \,. \tag{B.20}$$

PROOF It is based on the Ito's formula applied to $Y = f(X)$.

$$dY = \sum_i \frac{\partial f}{\partial x_i}(X) dX_i + 1/2 \sum_{i,j \leq d} \frac{\partial^2 f}{\partial x_i \partial x_j}(X) d\langle X_i^c, X_j^c \rangle \,,$$

from which, denoting by $\mathbb{E}^x[.]$ the conditional expectation with respect to the event $X_0 = x$, it can be deduced that

$$\mathbb{E}^x[f(X_t)] = f(x)$$

$$+\mathbb{E}^x \left[\int_0^t \left(\frac{\partial f}{\partial t}(X) + \sum_{i \leq d} b_i(x) \frac{\partial f}{\partial x_i}(X) + 1/2 \sum_{i,j \leq d} (\sigma\sigma')_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(X) \right) ds \right] .$$

Then, the proposition is established by applying the definition of A

$$Af(x) = \lim_{t \to 0} \frac{1}{t} \{\mathbb{E}[f(X_t) | X_0 = x] - f(x)\}.$$

$\square$

Example B.4 **Generators**

1) The $d-$dimensional Brownian motion, which of course is the solution of

$$dX_t = dW_t,$$

has a generator A given by, for any function f in $C_0^2(\mathbb{R}^d)$,

$$Af = 1/2 \sum_i \frac{\partial^2 f}{\partial x_i^2}(X).$$

Thus $A = 1/2\,\Delta$ where Δ is the Laplace operator.

2) The Ornstein-Uhlenbeck process

$$dX_t = k(m - X_t)dt + \sigma dW_t,$$

has a generator A given by

$$Af(x) = k(m-x)\frac{\partial f}{\partial x} + 1/2\sigma^2 \frac{\partial^2 f}{\partial x^2}\,.$$

$\square$

Other standard concepts for Markov processes are the following ones (see *e.g.* Oksendal [401]):

1) The Dynkin formula.

PROPOSITION B.4
Assume that f is in $C_0^2(\mathbb{R}^d)$. If T is a stopping time with $\mathbb{E}[T|X_0 = x] < \infty$ then

$$\mathbb{E}[f(X_T)|X_0 = x] = f(x) + \mathbb{E}[\int_0^T Af(X_s)ds|X_0 = x]\,. \tag{B.21}$$

For example, if T is the first exit time of a bounded set, then $\mathbb{E}[T|X_0 = x] < \infty$ and the above formula is valid.

Consider for example the d-dimensional Brownian motion W starting at $x_0 \in \mathbb{R}^d$. Assume that $|a| < R$. Then, by applying the Dynkin formula, the expected value of the first exit time T_R of the Brownian motion from the ball

$$\{x \in \mathbb{R}^d;\ |x| < R\}$$

is given by

$$\mathbb{E}[T_R|W_0 = x_0] = 1/d\,(R^2 - |x_0|^2)\,.$$

2) The Kolmogorov's backward equation.

Consider an Ito diffusion X in $\mathbb{R}^d$ with a generator A. If f is a function in $C_0^2(\mathbb{R}^d)$, then by using the Dynkin formula with $T = t$, it is obvious that

$$u(t,x) = \mathbb{E}[f(X_T)|X_0 = x]$$

is differentiable with respect to t with a differential given by

$$\frac{\partial u}{\partial t} = \mathbb{E}^x[Af(X_T)] \,.$$

Therefore, we have the following result:

PROPOSITION B.5 *Solution of the backward equation*

1) If f is in $C_0^2(\mathbb{R}^d)$, then $u(t,.)$ is in $\mathcal{D}(A)$ for each t and

$$\begin{array}{ll} \frac{\partial u}{\partial t} = Au, & t>0,\, x\in\mathbb{R}^d, \\ u(0,x) = f(x), & x\in\mathbb{R}^d \,. \end{array} \tag{B.22}$$

2) Let $C^{1,2}(\mathbb{R}^d)$ be the set of all functions continuous with respect to t and twice continuously differentiable with respect to x. If a function $w(.,.)$ is bounded and in $C^{1,2}(\mathbb{R}^d)$ and satisfies (B.22), then $w = u$.

3) The Feynman-Kac formula.
This is a generalization of the Kolmogorov's backward equation.

PROPOSITION B.6
If f is in $C_0^2(\mathbb{R}^d)$ and g is continuous on $\mathbb{R}^d$ and lower bounded. Then $v(t,.)$ defined by

$$v(t,x) = \mathbb{E}^x[exp\left(-\int_0^t g(X_s)ds\right) f(X_t)]$$

is the solution of

$$\begin{array}{ll} \frac{\partial v}{\partial t} = Av - gv, & t>0,\, x\in\mathbb{R}^d, \\ v(0,x) = f(x), & x\in\mathbb{R}^d \,. \end{array} \tag{B.23}$$

2) Let $C^{1,2}(\mathbb{R}^d)$ be the set of all functions continuous with respect to t and twice continuously differentiable with respect to x. If a function $w(.,.)$ is in $C^{1,2}(\mathbb{R}^d)$, is bounded on $K\times\mathbb{R}^d$ for each compact subset K in $\mathbb{R}$, and satisfies (B.23) then $w = v$.

An application of this is the determination of the generator of a killed diffusion. Consider an Ito process X solution of

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \,.$$

Its generator is given by

$$Af(x) = \sum_{i\le d} b_i(x)\frac{\partial f}{\partial x_i} + 1/2\sum_{i,j\le d}(\sigma\sigma')_{i,j}(x)\frac{\partial^2 f}{\partial x_i\partial x_j} \,.$$

Consider now the process $\tilde{X}$ which is the process X killed at the random time T :

$$\tilde{X}_t = X_t, \text{ if } t < T \,,$$

and $\tilde{X}_t$ is undefined if $t \geq T$. Denote the function $c(x)$, which is the killing rate defined by

$$c(x) = \lim_{t \downarrow 0} 1/t\, \mathbb{P}[X \text{ is killed in the time interval } (0,t) | X_0 = x] \,.$$

Then, $\tilde{X}$ is also a strong Markov process and

$$\mathbb{E}^x[f(\tilde{X}_t)] = \mathbb{E}^x[f(X_t)1_{t<T}] = \mathbb{E}^x[f(X_t)exp(-\int_0^t c(X_s)ds)] \,,$$

for all bounded continuous functions f on $\mathbb{R}^d$. Thus, the generator $\tilde{A}$ of $\tilde{X}$ is given by

$$\tilde{A}f(x) = \sum_{i \leq d} b_i(x) \frac{\partial f}{\partial x_i} + 1/2 \sum_{i,j \leq d} (\sigma\sigma')_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j} - c(x)f(x) \,.$$

As it can be seen, the use of the generators may lead to the resolution of partial differential equations (PDE), in particular when calculating option prices. Consider the following pricing problem:
Let A be the operator

$$Af(t,x) = \frac{\partial f}{\partial t}(t,x) + b(t,x)\frac{\partial f}{\partial x}(t,x) + 1/2\sigma^2(t,x)\frac{\partial^2 f}{\partial x^2}(t,x) \,,$$

where b and σ satisfy the assumptions of Theorem (B.3).
For a given function g (the payoff of a European option, for example), find the solutions f of the next parabolic equation ($f(t, X_t)$ will be the price of the option at any time t according to the value of the stock X_t at time t):

$$\begin{aligned} Af(t,x) &= rf(t,x), \mathcal{F}orallt \in [0,T], \forall x \in \mathbb{R}^d, \\ f(T,x) &= g(x). \end{aligned} \tag{B.24}$$

Consider $X^{x,t}$ the Ito process defined by, for all $u \geq t$,

$$X_u^{x,t} = X_t^{x,t} + \int_t^u b(s, X_s^{x,t})ds + \int_t^u \sigma(s, X_s^{x,t})dW_s,$$

with the initial condition $X_t^{x,t} = x$. Then, if f is the solution of (B.24), applying the Ito formula, f satisfies

$$f(u, X_u^{x,t})e^{-ru} = e^{-rt}f(t,x) + \int_t^u e^{-rs}\frac{\partial f}{\partial x}(s, X_s^{x,t})\sigma(s, X_s^{x,t})dW_s.$$

If the previous stochastic integral is a martingale (with a mild integrability assumption on $\frac{\partial f}{\partial x}$), then it can be deduced that

$$f(t,x) = \mathbb{E}[e^{-r(T-t)}g(X_T^{x,t})] = \mathbb{E}[e^{-r(T-t)}g(X_T)|X_t = x].$$

Example B.5 **Basic models**
1) Black and Scholes model - The well-known equation is

$$\begin{aligned} Af(t,x) &= \tfrac{\partial f}{\partial t} + rx\tfrac{\partial f}{\partial x} + 1/2\sigma^2\tfrac{\partial^2 f}{\partial x^2} = rf(t,x)\ , \\ f(T,x) &= g(x) = (x-K)^+\ . \end{aligned} \tag{B.25}$$

So

$$f(t,x) = \mathbb{E}[e^{-r(T-t)}(X_T - K)^+|X_t = x]\ .$$

2) Cox-Ingersoll-Ross model - This process is usually used in the term structure modeling.
Thus, it is important to calculate such expectations as $\mathbb{E}\left[exp(-\int_s^t r_u du|\mathcal{F}_s)\right]$ when the spot rate r is given by:

$$dr_t = a(b - r_t)dt + \rho\sqrt{r_t}dW_t.$$

Using the Markov property, it remains to calculate $\mathbb{E}\left[exp(-\int_0^t r_u du)\right]$. For this, consider the solution $r_u^{x,t}$ of

$$dr_s^{x,t} = a(b - r_s^{x,t})ds + \rho\sqrt{r_s^{x,t}}dW_s,\ r_t^{x,t} = x\ .$$

Introduce :

$$f(t,x) = \mathbb{E}\left[exp\left(-\int_t^T r_u^{x,t}du\right)\right],\ f(T,x) = 1\ .$$

Then, f is solution of the PDE :

$$\frac{\partial f}{\partial t} + a(b-x)\frac{\partial f}{\partial x} + 1/2\rho^2\frac{\partial^2 f}{\partial x^2} = xf\ .$$

Now, results concerning PDE are applied to determine the solution which is given by

$$f(t,x) = \Phi(T-t)exp(-r(t)\Psi(T-t))\ ,$$

with $\Phi(s) = \left(\frac{2\gamma e^{(\gamma+a)s/2}}{(\gamma+a)(e^{\gamma s}-1)+2\gamma}\right)^{\frac{2ab}{\rho^2}}$, $\gamma^2 = a^2 + 2\rho^2$, $\Psi(s) = \frac{2(e^{\gamma s}-1)}{(\gamma+a)(e^{\gamma s}-1)+2\gamma}$. □

Ito diffusions are used in most financial models. Nevertheless, other semimartingales can be introduced instead of the Brownian motion; other Lévy processes for example (see Cont and Tankov [128]). So, it is interesting to study stochastic differential equations driven by more general semimartingales (see Protter [419]).

References

[1] Acerbi, C., (2002): Spectral measures of risk: a coherent representation of subjective risk aversion, *Journal of Banking and Finance*, 26, 1505-1518.

[2] Acerbi, C., (2004): Coherent representations of subjective risk-aversion, In: G. Szegö (ed.), *Risk Measures for the 21st Century*, Wiley: New York.

[3] Acerbi, C., Nordio, C., and Sirtori, C., (2001): Expected shortfall as a tool for financial risk management, Working paper, http://www.gloriamundi.org.

[4] Acerbi, C. and Prospero, S., (2002): Portfolio optimization with spectral measures of risk, Abaxbank.

[5] Acerbi, C. and Tasche, D., (2002): On the coherence of expected shortfall, *Journal of Banking and Finance*, 26, 1487-1503.

[6] Adcock, C.J. and Meade, N., (1994): A simple algorithm to incorporate transactions costs in quadratic optimization, *European Journal of Operational Research*, 79, 85-94.

[7] Agarwal, V. and Naik, N.Y., (2000a): Multi-period performance persistence analysis of hedge funds, *Journal of Financial and Quantitative Analysis*, 35, 327-342.

[8] Agarwal, V. and Naik, N.Y., (2000b): On taking the alternative route: risks, rewards style and performance persistence of hedge funds, *Journal of Alternative Investment*, 2, 6-23.

[9] Agarwal, V. and Naik, N.Y., (2000c): Performance evaluation of hedge funds with option-based and buy-and-hold strategies, working paper, London Business School.

[10] Agarwal, V. and Naik, N.Y., (2004): Risks and portfolio decisions involving hedge funds, *Review of Financial Studies,* 17, 63-98.

[11] Akaike, H., (1976): *Canonical correlation analysis of time series and the use of an information criterion*, System identification: advances and case studies, R.K. Mehra and D.G. Lainiotis, eds., Academic Press: New York, 27-96.

[12] Akian M., Menaldi J., and Sulem A., (1996): On an investment-consumption model with transaction costs, *SIAM Journal Control and Optimization*, 34, 329-364.

[13] Akian M., Sulem A., and Taksar M., (2001): Dynamic optimization of long term growth rate for a portfolio with transaction costs: the logarithmic utility case, *Mathematical Finance*, 11, 153-188.

[14] Alexander, C., (2002): *Market Models*, Wiley: New York.

[15] Alexander, C. and Dimitriu, A., (2005): Indexing and statistical arbitrage, *Journal of Portfolio Management*, 32, 50-63.

[16] Alexander, C. and Tasche, D., (2002): On the covariance matrices used in Value-at-Visk models, *Journal of Derivatives*, 4, 50-62.

[17] Alexander, G.J., (1993): Short selling and efficient sets, *Journal of Finance,* 48, 1497-1506.

[18] Alexander, G. and Baptista, A., (2002): Economic implications of using a mean-VaR model for portfolio selection: a comparison with mean-variance analysis, *Journal of Economic Dynamics and Control,* 34, 329-364.

[19] Allais, M., (1953): Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine, *Econometrica*, 21, 503-546.

[20] Allais, M., (1979): The so-called Allais paradox and rational decisions under uncertainty, in *Expected Utility Hypothesis and the Allais Paradox*, Allais and Hagen (1979) (eds.), Dordrecht: D. Reidel.

[21] Althofer, I. and Koschnick, K.U., (1991): On the convergence of threshold accepting, *Applied Mathematical Optimization*, 24, 183-195.

[22] Amenc, N. and Le Sourd, V., (2003): *Portfolio Theory and Performance Analysis*, Wiley Finance, Wiley.

[23] Amenc, N. and Martellini, L., (2002): Portfolio optimization and hedge fund style allocation decisions, *Journal of Alternative Investments*, 5, 7-20.

[24] Amenc N., Bonnet S., Henry G., Martellini L., and Weytens, A., (2004): *La Gestion Alternative*, Economica: Paris.

[25] Andersson, F., Mausser, H., Rosen, D., and Uryasev, S., (2001): Credit risk optimization with conditional Value-at-Risk criterion, *Mathematical Programming*, 89, 273-291.

[26] Anscombe, F.J. and Aumann, R.J., (1963): A definition of subjective probability, *Annals of Mathematical Statistics*, 34, 199-205.

[27] Avouyi-Dovi, S., Morin, S., and Neto, D., (2004): Optimal asset allocation with Omega function, working paper Banque de France, Paris, France.

[28] Arrow, K.J., (1953): Le rôle des valeurs boursières pour la répartition la meilleure des risques, Econométrie, Paris: CNRS (translated as Arrow, K. J., (1964): The role of securities in the optimal allocation of risk-bearing. *Review of Economics Studies*, 31, 91-96.)

[29] Arrow, K.J., (1965): *Aspects of the Theory of Risk-Bearing*, Helsinki: Yrj Hahnsson Foundation.

[30] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D., (1997): Thinking coherently, *Risk*, 10, 68-71.

[31] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D., (1999): Coherent measures of risk, *Mathematical Finance*, 9, 203-228.

[32] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., and Hyejin Ku, (2002): Coherent multiperiod risk measurement, Working paper, ETH Zürich.

[33] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., and Hyejin Ku, (2003): Coherent multiperiod risk adjusted values and Bellman's principle, Working paper, ETH Zürich.

[34] Athayde, G., (2001): Building a mean-downside risk portfolio frontier, in Sortino F. and S. Satchell, *Managing Downside Risk in Financial Markets: Theory, Practice and Implementation*, Butterworth Heinemann.

[35] Athayde, G. and Flôres, R. Jr., (2004): Finding a maximum skewness portfolio: a general solution to three-moments portfolio choice, *Journal of Economic Dynamics and Control*, 28, 1335-1352.

[36] Bacmann, J.-F. and Scholz, S., (2003): Alternative performance measures for hedge funds, *AIMA Journal*, June.

[37] Bajeux, I., (1989): Gestion de portefeuille dans un modèle binomial, *Annales d'Economie et de Statistiques*, 13, 50-76.

[38] Bajeux-Besnainou, I., Jordan, J.V., and Portait, R., (2001): An asset allocation puzzle: comment, *American Economic Review*, *91*, 1170-1179.

[39] Bajeux-Besnainou, I., Jordan, J.V., and Portait, R., (2003): Dynamic asset allocation for stocks, bonds and cash, *Journal of Business*, 76, 263-287.

[40] Bamberg, G., and Wagner, N., (2000): Equity index replication with standard and robust regression estimators, *OR Spectrum*, 22, 525-543.

[41] Baquero, G., Horst, J., and Verbeek, M., (2005): Survival, look-ahead bias, and persistence in hedge fund performance, *Journal of Financial and Quantitative Analysis,* 40, 493-517.

[42] Barra, (1996): The Barra E2 multiple-factor model, Barra web site.

[43] Barrieu, P. and El Karoui, N., (2002): Inf-convolution of risk measures and optimal risk transfer, *Finance and Stochastics*, 9, 269-298.

[44] Barrieu, P. and El Karoui, N., (2005): Pricing, hedging and optimally designing derivatives via minimization of risk measures, L.S.E., Statistics Department, London, and CMAP, Ecole Polytechnique, France.

[45] Bartlett, M., (1951): An inverse matrix adjustment arising in discriminant analysis, *Annals of Mathematical Statistics*, 22, 107-111.

[46] Basak, S., (1995): A general equilibrium model of portfolio insurance, *Review of Financial Studies*, 8, 1059-1090.

[47] Basak, S., and Shapiro, A., (2001): Value-at-Risk-based risk management: optimal policies and asset prices, *Review of Financial Studies*, 14, 371-405.

[48] Basak, S., (2002): A comparative study of portfolio insurance, *Journal of Economic Dynamics and Control*, 26, 1217-1241.

[49] Bassi, F., Embrechts, P., and Kafetzaki, M., (1997): Risk management and quantile estimation, in: Adler, R., Feldman, R., and Taqqu, M. (Eds.), Practical Guide to Heavy Tails, Birkhäuser: Boston.

[50] Battocchio, P., Menoncin, F., and Scaillet, O., (2005): Optimal asset allocation for pension funds under mortality risk during the accumulation and decumulation phases, forthcoming in *Annals of Operations Research*.

[51] Bauer, R.J., (1994): *Genetic Algorithms and Investment Strategies*, Wiley.

[52] Beasley, J.E., Meade N., and Chang, T.-J., (2003): An evolutionary heuristic for the index tracking problem, *European Journal of Operational Research*, 148, 621-643.

[53] Benzion, U. and Yagil, J., (2003): Portfolio composition choice: a behavioral approach, *Journal of Behavioral Finance*, 4, 85-95.

[54] Bernoulli, D., (1738): Exposition of a new theory on the measurement of risk, *Comentarii Academiae Scientiarium Imperialis Petropolitanae*, as reprinted in 1954, *Econometrica*, 22, 23-26.

[55] Bertsimas, D., Lauprete, G.J., and Samarov, A., (2004): Shortfall as a risk measure: properties, optimization and applications, *Journal of Economic Dynamics and Control*, 28, 1353-1381.

[56] Bertrand, P., (2005): A note on portfolio performance attribution: taking risk into account, *Journal of Asset Management*, 5, 428-437.

[57] Bertrand, P., Lesne, J.-P., and Prigent, J.-L., (2001): Gestion de portefeuille avec garantie: l'allocation optimale en actifs dérivés, *Finance*, 22, 7-35.

[58] Bertrand, P. and Prigent, J.-L., (2002): Portfolio Insurance: the extreme value approach of the CPPI method, *Finance*, 23, 69-86.

[59] Bertrand, P. and Prigent, J.-L., (2003): Portfolio insurance strategies: a comparison of standard methods when the volatility of the stock is stochastic, *Int. J. Business*, 8, 461-472.

[60] Bertrand, P. and Prigent, J.-L., (2005): Portfolio insurance strategies: OBPI versus CPPI, *Finance*, 26, 5-32.

[61] Bertrand, P., Prigent, J.-L., and Sobotka, R., (2001): Optimisation de portefeuille sous contrainte de variance de la tracking-error, *Banque & Marchés*, 13, 1826-1836.

[62] Bertrand, P. and Rousseau, P., (2005): L'attribution de performance: problématique, limite et extension, *Revue Française de Gestion.*

[63] Bertsekas, D.P., (1987): *Dynamic Programming*, Prentice-Hall.

[64] Bertsekas, D.P., (2005): *Dynamic Programming and Optimal Control*, vol I and II, Athena Scientific.

[65] Best, M.J. and Grauer, R., (1991): On the sensitivity of mean-variance efficient portfolios to changes in asset means: some analytical and computational results, *Review of Financial Studies*, 4, 315-342.

[66] Best, M.J. and Grauer, R., (1992): Positively weighted minimum-variance portfolios and the structure of asset expected returns, *Journal of Financial and Quantitative Analysis*, 27, 513-537.

[67] Best, M.J. and Hlouskova, J., (2000): The efficient frontier for bounded assets, *Mathematical Methods of Operations Research*, 52, 195-212.

[68] Best, M.J., and Hlouskova, J., (2003): Portfolio selection and transactions costs, *Computational Optimization and Applications*, 24, 95-116.

[69] Best, M.J. and Kale, J., (2000): Quadratic programming for large-scale portfolio optimization, in *Financial Services Information Systems*, ed. Jessica Keyes, CRC Press, 513-529.

[70] Bielecki, T.R., and Pliska, S. R., (2000): Risk-sensitive dynamic asset management in the presence of transaction costs, *Finance and Stochastics*, 4, 1-33.

[71] Bion-Nadal, J., (2003): Conditional risk measure and robust representation of convex conditional risk measures, C.M.A.P., Ecole Polytechnique, Palaiseau, France.

[72] Bixby, R.E., Gregory, J. W., Lustig, I.J., Marsten, R.E., and Shanno, D.F., (1992): Very large scale linear programming: a case study in combining interior point and simplex methods, *Operations Research*, 40, 885-897.

[73] Black, F., (1972): Capital market equilibrium with restricted borrowing, *Journal of Business,* 45, 444-455.

[74] Black, F. and Litterman, R., (1990): Asset allocation: combining investor views with market equilibrium, Goldman Sachs Fixed Income Research.

[75] Black, F. and Rouhani, R., (1989): Constant proportion portfolio insurance and the synthetic put option: a comparison, in *Institutional investor focus on investment management*, edited by Frank J. Fabozzi, Cambridge, Mass. : Ballinger, 695-708.

[76] Black, F. and Jones, R., (1987): Simplying portfolio insurance, *Journal of Portfolio Management*, 48-51.

[77] Black, F. and Perold, A.R., (1992): Theory of constant proportion portfolio insurance, *Journal of Economic Dynamics and Control*, 16, 403-426.

[78] Black, F. and Scholes, M., (1973): The pricing of options and corporate liabilities portfolio insurance, *Journal of Political Economy*, 81, 637-654.

[79] Blanchet-Scalliet, C., El Karoui, N., Jeanblanc, M., and Martellini, L., (2001): Optimal investment and consumption decisions when time-horizon is uncertain, working paper, University of Evry, France.

[80] Bollerslev, T., Engle, R., and Nelson, D., (1986): Arch Models, in: R. F. Engle, and D. McFadden (eds.), Handbook of Econometrics, Elsevier, Edition 1, volume 4, chapter 49, pages 2959-3038.

[81] Bonnet, A. and Nagot, I., (2005): Methodology of measuring performance in alternative investment, Working paper, University Paris 1, France.

[82] Bookstaber, R. and Langsam, J.A., (2000): Portfolio insurance trading rules, *Journal of Futures Markets*, 8, 15-31.

[83] Bookstaber, R. and Roger, C., (1985): Problems in evaluating the performance of portfolios with options, *Financial Analysts Journal*, 41, 48-62.

[84] Bouchard, B. and Pham, H., (2004): Wealth-path dependent utility maximization in incomplete markets, *Finance and Stochastics*, 8, 579-603.

[85] Borodine, A.N. and Salminen, P., (2002): *Handbook of Brownian Motion - Facts and Formulae*, Birkhäuser.

[86] Boyd, S. and Vandenberghe, L., (2004): *Convex Optimization*, Cambridge University Press.

[87] Brennan, M., (1970): Taxes, market valuation and corporate financial policy, *National Tax Journal*, 25, 417-427.

[88] Brennan, M. and Solanki, R., (1981): Optimal portfolio insurance, *Journal of Financial and Quantitative Analysis*, 16, 279-300.

[89] Brennan, M. and Xia, Y., (2000): Stochastic interest rates and the Bond-Stock mix, *European Finance Review*, 4, 197-210.

[90] Brinson, G.P., Hood, L.R., and Beebower, G.L. (1986): Determinants of portfolio performance, *Financial Analysts Journal,* 39-44.

[91] Brooks, C. and Kat, H., (2002): The statistical properties of hedge fund index returns and their implications for investors, *Journal of Alternative Investments*, 5, 26-44.

[92] Brown, W., Durbin, J., and Evans, J.M., (1975): Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society*, 37, 149-192.

[93] Brown, S.J., Goetzmann, W., Ibbotson, R.G., and Ross, S.A., (1992): Survivorship bias in performance studies, *Review of Financial Studies*, 5, 553-580.

[94] Browne, S., (1999a): Beating a moving target: optimal portfolio strategies for outperforming a stochastic benchmark, *Finance and Stochastics*, 3, 275-294.

[95] Browne, S., (1999b): Reaching goals by a dealine: digital options and continuous-time active portfolio management, *Advances in Applied Probability*, 31, 551-577.

[96] Buckley, I.R.C., and Korn, R., (1998): Optimal index tracking under transaction costs and impulse control, *International Journal of Theoretical and Applied Finance*, 1, 315-330.

[97] Cadenillas, A., (2000): Consumption-investment problems with transaction costs: survey and open problems, *Mathematical Methods Operations Research*, 51, 43-68.

[98] Cadenillas, A., and Pliska, S., (1999): Optimal trading of a security when there are taxes and transaction costs, *Finance and Stochastics,* 2, 137-165.

[99] Cadenillas, A. and Sethi, S.P., (1997): The consumption investment problems with subsistence consumption, bankruptcy and random market coefficients, *Journal of Optimization Theory and Applications*, 93.

[100] Caglayan, M. and Edwards, F., (2001): Hedge fund performance and manager skill, *Journal of Futures Markets,* 21, 1003-1028.

[101] Cai, X., Teo, K. L., Yang, X. Q., and Zhou, X.Y., (2000): Portfolio optimization under a minimax rule, *Management Science*, 46, 957-972.

[102] Campbell, J., Cocco, J., Gomes, F., Maenhout, P.J., and Viceira, L., (2001): Stock market mean reversion and the optimal equity allocation of a long-lived investor, *European Finance Review*, 5,269-292.

[103] Campbell, J. Y. and Viceira, L., (2002): *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Oxford University Press: Oxford.

[104] Canner, N., Mankiw, N. G., and Weil, D.N., (1997): An asset allocation puzzle, *American Economic Review*, 87, 181-191.

[105] Carhart, M.M., (1997): On persistence in mutual fund performance, *Journal of Finance,* 52, 57-82.

[106] Carr, P. and Madan, D., (2001): Optimal positioning in derivative securities, *Quantitative Finance*, 1, 19-37.

[107] Cascon, A., Keating, C., and Shadwick, F., (2002): The mathematics of the Omega measure, working paper, The Finance Development Center.

[108] Cass, D. and Stiglitz, J.E., (1970): The structure of investor preferences and asset returns, and separability in portfolio allocation: a contribution to the pure theory of mutual funds, *Journal of Economic Theory*, 2, 122-160.

[109] Cerf, R., (1994): *Une théorie asymptotique des algorithmes génétiques*, Phd Thesis, University of Montpellier II, France.

[110] Cesari, L., (1983): *Optimization Theory and Applications: Problems with Ordinary Equations*, Springer: New York.

[111] Chabaane, A., Laurent, J.-P., Malevergne, Y., and Turpin, F., (2003): Alternative risk measures for alternative investments, to appear in *Journal of Risk*.

[112] Chamberlain, G., (1983): A characterization of the distributions that imply mean-variance utility functions, *Journal of Economic Theory*, 29, 185-201.

[113] Chan, L.K.C., Karceski, J., and Lakonishok, J., (1999): On portfolio optimization: forecasting covariances and choosing the risk model, *Review of Financial Studies*, 12, 937-974.

[114] Chan, K., Karolyi, G., Longstaff, F., and Sanders, A., (1992): An empirical comparison of alternative models of the short term interest rate, *Journal of Finance*, 47, 1209-1227.

[115] Chateauneuf, A., Cohen, M., and Meilijson, I., (1997): New tools to better model behavior under risk and uncertainty: an overview, *Finance*, 18, 25-46.

[116] Chekhlov, A., Uryasev, S., and Zabarankin, M., (2003): Portfolio optimization with drawdown constraints, in *Asset and Liability Management Tools*, (B. Scherer ed.), Risk Books, London, 263-278.

[117] Cheridito, P., Delbaen, F., and Kupper, M., (2004): Coherent and convex risk measures for bounded cadlag processes, *Stochastic processes and their applications*, 112, 1-22.

[118] Cheridito, P., Delbaen, F., and Kupper, M., (2005): Coherent and convex risk measures for unbounded cadlag processes, *Finance and Stochastics*, 9, 369-387.

[119] Chew, S.-H., (1983): A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox, *Econometrica*, 51, 1065-1092.

[120] Chew, S.-H. and MacCrimmon, K.R., (1979): Alpha-nu choice theory: a generalization of expected utility theory, Working Paper No. 669, University of British Columbia, Vancouver.

[121] Chew, S., Karni, E., and Safra, Z., (1987): Risk aversion in the theory of expected utility with rank dependent preferences, *Journal of Economic Theory*, 42, 370-381.

[122] Chopra, V. K. and Ziemba, W.T., (1993): The effect of errors in means, variances and covariances on optimal portfolio choice, *Journal of Portfolio Management*, 19, 6-11.

[123] Choquet, G., (1953): Théorie des capacités, *Annales Institut Fourier*, 5, 131-295.

[124] Clarke, R.C., Krase, S., and Statman, M. (1994). Tracking errors, regret, and tactical asset allocation, *Journal of Portfolio Management,* 20, 16-24.

[125] Cohen, M. and Tallon, J.-M., (2000): Décision dans le risque et l'incertain : l'apport des modèles non-additifs, *Revue d'Economie Politique*, 110.

[126] Connor, G. and Leland, H., (1995): Cash management for index tracking, *Financial Analysts Journal*, 75-80.

[127] Consiglio, A. and Zenios, S., (1999): Designing portfolios of financial products via integrated simulation and optimization models, *Operations Research*, 47, 195-208.

[128] Cont, R. and Tankov, P., (2004): *Financial Modelling with Jump Processes*, Chapman & Hall: Boca Raton.

[129] Coquet, F., Hu, Y., Mémin, J., and Peng, S., (2002): Filtration consistent nonlinear expectations and related g-expectations, *Probability Theory and Related Fields*, 123, 381-411.

[130] Cornell, B., (1979): Asymmetric information and portfolio performance measurement, *Journal of Financial Economics*, 7, 381-390.

[131] Cornish, E.A. and Fisher, R.A., (1937): Moments and cumulants in the specification of distributions, *Review of the International Statistical Institute*, 307-330.

[132] Cox, J. and Huang, C.-F., (1989): Optimal consumption and portfolio choices when asset prices follow a diffusion process, *Journal of Economic Theory*, 49, 33-83.

[133] Cox, J. and Huang, C.-F., (1991): A variational problem arising in financial economics, *Journal of Mathematical Economics*, 20, 465-487.

[134] Cuoco, D., (1997): Optimal consumption and equilibrium prices with portfolio constraints and stochastic income, *Journal of Economic Theory*, 72, 33-73.

[135] Cuoco D., He, H., and Issaneko, S., (2001): Optimal dynamic trading strategies with risk limits, Working paper, Wharton School.

[136] Cvitanic, J., Goukasian, L., and Zapatero, F., (2003): Monte Carlo computation of optimal portfolios in complete markets, *Journal of Economic Dynamics and Control*, 27, 971-986.

[137] Cvitanic, J. and Karatzas, I., (1992): Convex duality in constrained portfolio optimization, *Annals of Applied Probability,* 2, 767-818.

[138] Cvitanic, J. and Karatzas, I., (1995): On portfolio optimization under drawdown constraints, *IMA Lecture Notes in Mathematics and Applications*, 65, 77-88.

[139] Cvitanic, J. and Karatzas, I., (1996): Hedging and portfolio optimization under transaction costs: a martingale approach, *Mathematical Finance,* 6, 133-165.

[140] Cvitanic, J. and Karatzas, I., (1999): On dynamic measures of risk. *Finance and Stochastics*, 3, 451-482.

[141] Cvitanic, J., Lazrak, A., Martellini, L., and Zapatero, F., (2003): Optimal allocation to hedge funds: an empirical analysis, *Quantitative Finance*, 3, 1-12.

[142] Cvitanic, J., Schachermayer, W., and Wang, H., (2001): Utility maximization in incomplete markets with random endowment, *Finance and Stochastics*, 5, 259-272.

[143] Daniel, K., Grinblatt, M., Titman, S., and Wermers, R., (1997): Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance*, 52, 1035-1058.

[144] Danielsson, J., Jorgensen, B.N., Samorodnitsky, G., Sarma, M., and de Vries, C.G., (2005): Subadditivity re-examined: the case for Value-at-Risk, Working paper, London School of Economics.

[145] Dantzig, G.B., (1998): *Linear Programming and Extensions*, Princeton University Press: Princeton, NJ.

[146] Dantzig, G.B. and Infanger, G., (1993): Multi-stage stochastic linear programs for portfolio optimization, *Annals of Operations Research*, 45, 59-76.

[147] Darinka, D. and Ruszczynski, A., (2006): Portfolio optimization with stochastic dominance constraints, *Journal of Banking and Finance*, 30, 433-451.

[148] Dash, S., Wagner, N., Brück, B., and Diller, C., (2004): Tracking an index with narrow baskets: efficiency, costs and tradeoffs involved in optimized portfolios, Standard and Poor's Report, Munich, New York, Paris, London.

[149] Davis, M. and Norman, A., (1990): Portfolio selection with transaction costs, *Mathematics of Operations Research*, 15, 676-713.

[150] Davis, M. (1993): *Markov Models and Optimization*, Chapman and Hall: London.

[151] De Bondt, W. and Wolff, F.P.C., (2004): Introduction to the special issue on behavorial finance, *Journal of Empirical Finance*, 11, 423-427.

[152] Debreu, G.,(1959): *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, Yale University Press: New Haven.

[153] Deelstra, G., Pham, H., and Touzi, N., (2001): Dual formulation of the utility maximization problem under transaction costs, *Annals of Applied Probability*, 11, 1353-1383.

[154] Delbaen, F., (2002): Coherent risk measures on general probability spaces, *Advances In Finance and Stochastics*, Essays in Honour of Dieter Sondermann, Springer: New York.

[155] Delbaen, F. and Schachermayer, W. (2006): *The mathematics of Arbitrage*, Springer: Berlin.

[156] Demange, G. and Rochet, J.-C., (1992): *Méthodes Mathématiques de la Finance*, Economica: Paris.

[157] Dembo, R.S. and Rosen, D., (1999): The practice of portfolio replication: a practical overview of forward and inverse problems, *Annals of Operations Research*, 85, 267-284.

[158] Dempster, M.A.H., ed., (1980): *Stochastic Programming*, Academic Press: London.

[159] Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R., and Laeven, R., (2005): Risk measurement with the equivalent utility principles, working paper, Université Catholique de Louvain, Belgium.

[160] De Palma, A., and Picard, N., (2003): Measuring risk tolerance, working paper, University of Cergy-Pontoise, THEMA, France.

[161] De Palma, A., and Prigent, J.-L., (2005): Standardized versus customized portfolio: a compensating variation approach, to appear in *Annals of Operations Research*.

[162] De Palma, A., Picard, N., and Prigent, J.-L., (2005): Utility invariance under a family of transformations, Working paper 2005, University of Cergy-Pontoise, Thema, France.

[163] Derigs, U. and Nickel, N., (2004): On a local-search heuristic for a class of tracking error minimization problems in portfolio management, *Annals of Operations Research*, 131, 45-77.

[164] Detemple, J. Garcia, R., and Rindishbacher, M., (2003): A Monte Carlo method for optimal portfolios, *Journal of Finance*, 58, 401-446.

[165] Detlefsen, K. and Scandolo, G., (2005): Conditional and dynamic convex risk measures. *Finance and Stochastics*, 9, 539-561.

[166] Dexter, A. S., Yu, J. N. W., and Ziemba, W.T., (1980): Portfolio selection in a lognormal market when the investor has a power utility: computational results, in *Stochastic Programming,* M.A.H. Dempster (Ed.), Academic Press, New-York, 507-523.

[167] Diamond, P. and Stiglitz, J. E., (1974): Increases in risk and in risk aversion, *Journal of Economic Theory*, 8, 337-360.

[168] Doherty, N. and Eeckhoudt, L., (1995):. Optimal insurance without expected utility : the dual theory and the linearity of insurance contracts, *Journal of Risk and Uncertainty*, 10, 157-179.

[169] Doukhan, P., (1994): *Mixing: Properties and Examples*, Lecture Notes in Statistics, Springer: New York.

[170] Dowd, K., (1998): *Beyond Value-at-Risk*: *The New Science of Risk Management*, Wiley: Chichester.

[171] Dueck, G. and Scheuer, T., (1990): Threshold accepting: a general purpose algorithm appearing superior to simulated annealing, *Journal of Computational Physics*, 90, 161-175.

[172] Dueck, G. and Winker, P., (1992): New concepts and algorithms for portfolio choice, *Applied Stochastic Models and Data Analysis*, 8, 159-178.

[173] Duffie, D. and Epstein, L.G., (1992): Stochastic differential utility, *Econometrica*, 60, 353-394.

[174] Duffie, D., Fleming, W., Soner, H.M. and Zariphopoulou, T., (1997): Hedging in incomplete markets with Hara utility, *Journal of Economic Dynamics and Control*, 21, 753-782.

[175] Duffie, D. and Huang, C.F., (1985): Implementing Arrow-Debreu equilibria by continuous trading a few long-lived securities, *Econometrica*, 53, 1337-1356.

[176] Duffie, D. and Pan, J. (1997): An overview of value at risk, *Journal of Derivatives*, 4, 7-49.

[177] Duffie, D. and Sun, T.S., (1990): Transactions costs and portfolio choice in a discrete-continuous time setting, *Journal of Economic Dynamics and Control*, 14, 35-51.

[178] Davidson, R. and Mac Kinnon, J., (1993): *Estimation and Inference in Econometrics*, Oxford University Press: New York.

[179] Dunis, C. and Ho, R., (2005): Cointegration portfolios of European equities for index tracking and market neutral strategies, *Journal of Asset Management*, 6, 33-52.

[180] Dickey, D.A. and Fuller, W.A., (1979): Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, 74, 427-431.

[181] Dybvig, P.H., (1984): Short sales restrictions and kinks on the mean variance frontier, *Journal of Finance*, 39, 239-244.

[182] Dybvig, P.H., (1988): Inefficient dynamic portfolio strategies or how to throw away a million dollars in the stock market, *Review of Financial Studies*, 1, 67-88.

[183] Dybvig, P.H. and Ingersoll, J.E., (1982): Mean-variance theory in complete markets, *Journal of Business*, 55, 233-251.

[184] Dybvig, P.H. and Ross, S.A., (1985): The analytics of performance measurement using a security market line, *Journal of Finance*, 40, 401-416.

[185] Eeckhoudt, L., (1997): La théorie duale des choix risqués et la diversification: quelques réflexions, *Finance*, 18, 77-84.

[186] Edgeworth, F.Y., (1908): *Papers relating to political economy*, 3 vols (London: MacMillan for the Royal Economic Society, 1925), I, 19-20, fn. 3.

[187] Ekeland, I. and Turnbull, T., (1983): *Infinite-Dimensional Optimization and Convexity*, Chicago lectures in Mathematics, the University of Chicago Press.

[188] El Karoui, N., (1981): *Les Aspects Probabilistes du Contrôle Stochastique*, Lecture Notes in Math 876, Springer, Berlin, New York.

[189] El Karoui, N. and Karatzas, I., (1994): Dynamic allocation problems in continuous time, *Annals of Applied Probability*, 4, 255-286.

[190] El Karoui, N., and Jeanblanc, M., (1998): Optimization of consumption with labor income, *Finance and Stochastics*, 2, 409-440.

[191] El Karoui, N., Jeanblanc, M., and Huang, S., (2002): Random horizon, working paper, University of Evry, France.

[192] El Karoui, N., Jeanblanc, M., and Lacoste, V., (2005): Optimal portfolio management with American capital guarantee, *Journal of Economic Dynamics and Control*, 29, 449-468.

[193] El Karoui, N. and Meziou A., (2006): Constrained optimization with respect to stochastic dominance: application to portfolio insurance, *Mathematical Finance*, 16, 103-117.

[194] El Karoui, N. and Quenez, M.-C., (1996): Non-linear pricing theory and backward stochastic differential equations in financial mathematics, (ed: W.J. Runggaldier), *Lecture Notes in Mathematics* 1656, Springer Verlag, 191-246.

[195] El Karoui, N., S. Peng, S., and Quenez, M.-C., (1997): Backward stochastic differential equations in finance, *Mathematical Finance*, 7, 1-71.

[196] Ellsberg, D., (1961): Risk, ambiguity and the Savage axioms, *Quarterly Journal of Economics*, 75, 643-669.

[197] Elton, E. and Gruber, M., (1973): Estimating the dependence structure of share prices-Implications for portfolio selection, *Journal of Finance*, 8, 1203-1232.

[198] Elton, E., Gruber, M., and Padberg, M., (1976): Simple criteria for optimal portfolio selection, *Journal of Finance*, 31, 1341-1357.

[199] Elton, E., Gruber, M., and Padberg, M., (1978): Optimal portfolio from simple ranking devices, *Journal of Portfolio Management*, 4, 15-19.

[200] Elton, E. and Gruber, M., (2002): *Modern Portfolio Theory and Investment Analysis*, 6th edition, Wiley: New York.

[201] Embrechts, P., Kluppelberg, C., and Mikosch, T., (1997): *Modelling Extremal Events for Finance and Insurance*, Springer: New York.

[202] Embrechts, P., Mc Neil, A., and Straumann, D., (1999): Correlation: pitfalls and alternatives, *Risk Magazine*, 69-71.

[203] Emmer, S., Klüppelberg, C., and Korn, R., (2001): Optimal portfolios with bounded Capital-at-Risk, *Mathematical Finance*, 11, 365-384.

[204] Engle, R.F., (1982): Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, *Econometrica*, 50, 987-1008.

[205] Engle, R.F. and Granger, C.W.J., (1987): Cointegration and error correction: representation, estimation and testing, *Econometrica*, 55, 251-276.

[206] Engle, R.F. and Kroner, K., (1995): Multivariate simultaneous GARCH, *Econometric Theory*, 11, 122-150.

[207] Engle, R. and Manganelli, S., (2004): Caviar: conditional autoregressive Value-at-Risk by regression quantiles, *Journal of Business & Economic Statistics*, 22, 367-381.

[208] Engle, R. and Ng, V., (1993): Time varying volatility and the dynamic behavior of the term structure, *Journal of Money, Credit and Banking*, 25, 336-349.

[209] Engle, R. and Sheppard, K., (2001): Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH, UCSD working paper, 2001-15.

[210] Epstein, L. G., (1985): Decreasing risk aversion and mean-variance analysis, *Econometrica*, 53, 945-962.

[211] Epstein, L.G. and Schneider, M., (2003): Recursive multiple-priors, *J. Econom. Theory*, 113, 1-31.

[212] Epstein, L.G. and Zin, S.E., (1989): Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework, *Econometrica*, 57, 937-969.

[213] Ested, T. and Kritzman, M., (1988): TIPP: insurance without complexity, *Journal of Portfolio Management*, 14, 38-42.

[214] Eun, C. and Resnick, B., (1992): Forecasting the correlation structure of share prices: a test of new models, *Journal of Banking and Finance*, 16, 643-656.

[215] Fabozzi, F., (1996): *Bond Markets Analysis and Strategies*, 3rd ed. Englewood Cliffs, NJ: Prentice Hall.

[216] Falk, M., (1985): Asymptotic normality of the kernel quantile estimator, *Annals of Statistics*, 13, 428-423.

[217] Fama, E.F., (1970): Efficient capital markets: a review of theory and empirical work, *Journal of Finance*, 25, 383-417.

[218] Fama, E.F., (1972): Components of investment performance, *Journal of Finance*, 27, 551-567.

[219] Fama, E.F. and French, K.R., (1993): Common risks factors in the returns on stocks and bonds, *Journal of Financial Economics*, 33, 389-406.

[220] Fang, K.-T., Kotz, S., and Ng, K.W., (1990): *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.

[221] Fermanian, J.D. and Scaillet, O., (2005): Sensitivity analysis of VaR and expected shortfall for portfolios under netting agreements, *J. Banking and Finance*, 29, 927-958.

[222] Ferson, W. E. and Schadt, R. W., (1996): Measuring fund strategy and performance in changing economic conditions, *Journal of Finance*, 51, 425-461.

[223] Fishburn, P.C., (1970): *Utility Theory for Decision-Making*, New York: Wiley.

[224] Fishburn, P.C., (1983): Transitive measurable utility, *Journal of Economic Theory*, 31, 293-317.

[225] Fishburn, P.C., (1988a): Expected utility: an anniversary and a new era, *Journal of Risk and Uncertainty*, (Historical archive), 1, 267-283.

[226] Fishburn, P.C., (1988a): *Non-linear Preference and Utility Theory*, Baltimore: Johns Hopkins University Press.

[227] Fischer, I., (1906): *The Nature of Capital and Income*, MacMillan: New York.

[228] Fischer, T., (2001): Examples of coherent risk measures depending on higher moments, *Insurance: Mathematics and Economics*, 32, 135-146.

[229] Fitzpatrick, B. and Fleming, W., (1991): Numerical methods for optimal investment-consumption models, *Math. Oper. Res.*, 16, 657-823.

[230] Fleming, W., (1978): Exit probabilities and optimal stochastic control, *Appl. Math. Optim.*, 4, 329-346.

[231] Fleming, W. and Rischel, R., (1975): *Deterministic and Stochastic Optimal Control*, Springer Verlag: Berlin.

[232] Fleming, W. and Soner, M., (1993): *Controlled Markov Processes and Viscosity Solution*, Springer Verlag: Berlin.

[233] Föllmer, H. and Kabanov, Y., (1998): Optional decomposition and Lagrange multipliers, *Finance and Stochastics*, 2, 69-81.

[234] Föllmer, H. and Leukert, P., (1999): Quantile hedging, *Finance and Stochastics*, 3, 251-273.

[235] Föllmer, H. and Leukert, P., (2000): Efficient hedging: cost versus shortfall risk, *Finance and Stochastics*, 4, 117-146.

[236] Föllmer, H. and Schied, A., (2002a): Convex measures of risk and trading constraints, *Finance and Stochastics*, 6, 429-448.

[237] Föllmer, H. and Schied, A. (2002b): Robust preferences and convex measures of risk, in K. Sandmann and P. J. Schönbucher, eds, *Advances in Finance and Stochastics*, Springer Verlag: Berlin, 39-56.

[238] Föllmer, H. and Schied, A., (2002c): *Stochastic Finance - An Introduction in Discrete Time*, Walter de Gruyter: Berlin.

[239] Fong, F., (1990): Portfolio construction: fixed income, in John L. Maginn and Donald L. Tuttle, eds., *Managing investment portfolios: a dynamic process*, 2nd ed. Boston: Warren, Gorham and Lamont.

[240] Frank, M. and Wolfe, P., (1956): An algorithm for quadratic programming, *Naval Research Logistics Quarterly*, 3, 149-154.

[241] Franks, E.C. (1992): Targeting excess-of-benchmark returns, *Journal of Portfolio Management* 18, 6-12.

[242] Friedman, M. and Savage, L., (1948): Utility analysis of choices involving risk, *Journal of Political Economics*, 56, 279-304.

[243] Frittelli, M. and Rosazza, G.E., (2004): Dynamic convex risk measures. In: G. Szegö (ed.), Risk Measures for the 21st Century, Wiley: New York.

[244] Frost, P. and Savarino, J., (1986): An empirical Bayes approach to efficient portfolio selection *Journal of Financial and Quantitative Analysis*, 21, 293-306.

[245] Frost, P. and Savarino, J., (1988): For better performance: constraint portfolio weights, *Journal of Portfolio Management*, 14, 29-34.

[246] Fung, W., and Hsieh, D.A., (1997a): Empirical characteristics of dynamic trading strategies: the case of hedge funds, *Review of Financial Studies*, 10, 275-302.

[247] Fung, W. and Hsieh, D.A., (1997b): Survivorship bias and investment style in the returns of CTA's, *Journal of Portfolio Management*, 24, 30-41.

[248] Fung, W. and Hsieh, D.A., (1997c): Hedge fund survival lifetimes, *Journal of Asset Management*, 3, 237-252.

[249] Fung, W. and Hsieh, D.A., (2001): The risk in hedge fund strategies: theory and evidence from trend followers, *Review of Financial Studies*, 14, 313-341.

[250] Gaivoronski, A. and Krylov, S., (2006): Decision support for benchmark tracking: dynamic rebalancing strategies, working paper, Department

of Industrial Economy and Technology Management, Norwegian University of Science and Technology, Norway.

[251] Gaivoronski, A. and Pflug, G., (2005): Value-at-Risk in portfolio optimization: properties and computational approach, *Journal of Risk*, 7, 1-31.

[252] Giacometti, R. and Lozza, S., (2004): Risk measures for asset allocation models, in the volume (eds. Szegö), Risk measures for the 21st century, Wiley, 69-87.

[253] Gilboa, I., (1987): Expected utility with purely subjective non-additive probabilities, *Journal of Mathematical Economics*, 16, 65-88.

[254] Gilboa, I. and Schmeidler, D. (1989): Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics*, 18, 141-153.

[255] Gilli, M. and Këllezi, E., (2002): The threshold accepting heuristic for index tracking, in *Financial Engineering, E-Commerce, and Supply Chain*, (Eds. P. Pardalos and V.K. Tsitsiringos), Kluwer Applied Optimization Series, 1-18.

[256] Glosten, L.R., Jagannathan, R., and Runkle, D., (1993): On the relation between the expected value and the volatility of the normal excess return on stocks, *Journal of Finance*, 48, 1779-1801.

[257] Goetzmann, W., Ingersoll, J.E., Spiegel, M., and Welch, I., (2002): Sharpening Sharpe ratios, Working Paper Yale School of Management.

[258] Gollier, C., (2001): *The Economics of Risk and Time*, The MIT Press: Cambridge, MA.

[259] Goovaerts, M.J., Kaasb, R., Dhaenea, J., and Tangb, Q., (2003): A unified approach to generate risk measures, Department of Quantitative Economics, University of Amsterdam.

[260] Gourieroux, C., (1997): *ARCH Models and Financial Applications*, Springer Series in Statistics, Springer: New york.

[261] Gourieroux, C., Laurent, J.-P., and Scaillet, O., (2000): Sensitivity analysis of values-at-risk, *Journal of Empirical Finance*, 7, 225-245.

[262] Granger, C.W.J., (1969): Investigating causal relations by econometric models and cross spectral methods, *Econometrica*, 37.

[263] Granger, C.W.J. and Joyeux, R., (1980): An introduction to longmemory time series models and fractional differencing, *Journal of Time Series Analysis*, 1, 15-29.

[264] Granger, C.W.J. and Newbold, P., (1974): Spurious regressions in econometrics, *Journal of Econometrics*, 2, 111-120.

[265] Green, R., (1986): Positively weighted portfolios on the minimum-variance frontier, *Journal of Finance*, 41, 1051-1068.

[266] Green, R. and Hollifield, B., (1992): When will mean-variance efficient portfolios be well diversified?, *Journal of Finance*, 47, 1785-1809.

[267] Greene, W.H., (2002): *Econometric Analysis*, 5th ed., Prentice Hall.

[268] Gregoriou, G., Hübner, G., and Papageorgiou, N., (2005): *Hedge funds: insights in performance measurement, risk analysis, and portfolio allocation*, Wiley.

[269] Grinblatt, M. and Titman, S., (1989a): Mutual fund performance: an analysis of quarterly portfolio holdings, *Journal of Business*, 62, 393-416.

[270] Grinblatt, M. and Titman, S., (1989b): Portfolio performance evaluation: old issues and new insights, *Review of Financial Studies*, 2, 393-421.

[271] Grinblatt, M. and Titman, S., (1993): Performance measurement without benchmarks: an examination of mutual fund returns, *Journal of Business*, 66, 47-68.

[272] Grinold, R.C., (1992): Are benchmark portfolios efficient?, *Journal of Portfolio Management*, 19, 34-40.

[273] Grinold, R.C., and Kahn, R., (2000): *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk*, 2nd edn., Irwin: Chicago.

[274] Grossman, S. and Vila, J.L., (1989): Portfolio insurance in complete markets: a note, *Journal of Business*, 62, 473-476.

[275] Grossman, S. and Zhou, J., (1993): Optimal investment strategies for controlling drawdowns, *Mathematical Finance*, 3, 241-276.

[276] Grossman, S. and Zhou, J., (1996): Equilibrium analysis of portfolio insurance, *Journal of Finance*, 51, 1379-1403.

[277] Hadar, J. and Russel, W., (1969): Rules for ordering uncertain prospects, *American Economic Review*, 59, 25-34.

[278] de Haan, L., Jansen, D.W., Koedijk, K., and de Vries C.G., (1994) : Safety first portfolio selection, extreme value theory and long run asset risks, *Extreme Value Theory and Applications*.

[279] Haar, S. and van Straalen, J.P., (2004): Managing tracking errors in a dynamic environment, *GARP Risk Review*, 16.

[280] Hamidi, B., Jurczenko, E., and Maillet, B., (2006): D'un multiple conditionnel en assurance de portefeuille, working paper, University of Paris 1.

[281] Hanoch, G. and Lévy, H., (1969): The efficiency analysis of choices involving risk, *Review of Economic Studies*, 36, 335-346.

[282] Harri, A. and Brorsen, B.W., (2004): Performance persistence and the source of returns for hedge funds, *Applied Financial Economics*, 14, 131-141.

[283] Harrel, F. and Davis, C., (1982): A new distribution free quantile estimation, *Biometrika*, 69, 635-640.

[284] Harrison, J.M. and Kreps, D., (1979): Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory*, 20, 381-408.

[285] Harrison, J.M. and Pliska, S., (1981): Martingales and stochastic integrals in the theory of continuous trading, *Stoch. Proc. Appl.*, 11, 215-260.

[286] Heath, D., (2000): Back to the future, Plenary lecture at the first World Congress of the Bachelier Society, June, Paris.

[287] He, H. and Pages, H.,(1993): Labor income, borrowing constraints, and equilibrium asset prices, *Economic Theory*, 3, 663-696.

[288] He, H. and Pearson, N.D., (1991a): Consumption and portfolio policies with incomplete markets and short-sale constraints: the infinite-dimensional case, *Journal of Economic Theory*, 54, 259-305.

[289] He, H. and Pearson, N.D., (1991b): Consumption and portfolio policies with incomplete markets and short-sale constraints: the finite-dimensional case, *Mathematical Finance*, 1, 1-10.

[290] Henriksson, R.D., and Merton, R.C., (1981): On market timing and investment performance II. Statistical procedures for evaluationg forecasting skills, *Journal of Business*, 54, 513-533.

[291] Herstein, I.N. and Milnor, J.,(1953) : An axiomatic approach to measurable utility, *Econometrica*, 21, 291-297.

[292] Herzog, F., Dondi, G., and Geering, H.P., (2004): Strategic asset allocation with factor models for returns and GARCH models for volatilities, Proceedings of the 4th IASTED International Conference on Modelling, Simulation, and Optimization - MSO 2004, pp. 59-65, Kauai, HI.

[293] Hicks, J., (1931): The theory of uncertainty and profit, *Economica*, 11, 170-189.

[294] Jacod, J. and Shiryaev, A.N., (1987): Limit Theorems for Stochastic Theorems, Grundlehren der Math. Wissenschaften, 288, Springer (new edition 2003).

[295] Jaffray, J.Y., (1989): Linear utility for belief functions, *Operations Research Letters*, 8, 107-112.

[296] Jeanblanc, M., Lacoste, V., and Roland, S., (2005): Utility maximization under partial information in jump-diffusion models, working paper, University of Evry, France.

[297] Jeanblanc, Lakner, P., and Kadam, A., (20014): Optimal bankruptcy time and consumption/investment policies on an infinite horizon with a continuous debt payment until bankruptcy, *Mathematics of Operations Research*, 29, 649-671.

[298] Jeanblanc, M. and Pontier, M., (1990): Optimal portfolio for a small investor in a market with discontinuous prices, *Applied mathematics and optimization*, 22, 287-310.

[299] Jegadeesh, N. and Titman, S., (1993): Returns to buying winners and selling losers: implications for stock market efficiency, *Journal of Finance*, 48, 65-91.

[300] Jensen, M.C., (1968): The performance of mutual funds in the period 1945-1964, *Journal of Finance*, 23, 389-416.

[301] Jensen, N.E., (1967): An introduction to Bernoullian utility theory, I: utility functions, *Swedish Journal of Economics*, 69, 163-83.

[302] Jensen, B.A. and Sorensen, C., (2001): Paying for minimum interest rate guarantees: who should compensate who? *European Financial Management*, 7, 183-211.

[303] Jia, J., and Dyer, S., (1996): A standard measure of risk and risk-value models, *Management Science*, 42, 1691-1705.

[304] Johansen, S., (1988): Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.

[305] Jobson, J. and Korkie, B., (1980): Estimation for Markowitz efficient portfolios, *Journal of the American Statistical Association*, 75, 544-554.

[306] Jobson, J. and Korkie, B., (1981): Putting Markowitz theory to work, *Journal of Portfolio Management*, 7, 70-74.

[307] Joe, H., (1997): *Multivariate Models and Dependence Concepts*, Chapmann & Hall: London.

[308] Jorion, P., (1996): Risk2: measuring the risk in Value-at-Tisk, *Financial Analysts Journal,* November/December, 47-56.

[309] Jorion, P., (1997): *Value at Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill: New York.

[310] Jorion, P., (2003): Portfolio optimization with tracking-error constraints, *Financial Analyst Journal*, 70-82.

[311] Jouini, E., Meddeb, M. and Nizar T., (2004): Vector-valued coherent risk measures, *Finance and Stochastics*, 8, 531-552.

[312] Jurczenko, E. and Maillet, B. (2004): The 4-CAPM: in between Asset Pricing and Asset Allocation, in: *Multi-moment Capital Asset Pricing Moment and Relatd Topics*, Adcock *et al.* (eds.), Springer, Chapter 3, forthcoming.

[313] Kahn, R.N. and Rudd, A., (1995): Does historical performance predict future performance?, Barra Newlsetter.

[314] Kahneman, D. and Tversky, A., (1979): Prospect theory: an analysis of decision under risk, *Econometrica*, 47, 263-291.

[315] Kallberg, J. G. and Ziemba, W.T., (1983): Comparison of alternative utility functions in portfolio selection problems, *Management Science*, 29, 1257-1276.

[316] Kan, Y. S. and Kibzun, A.I., (1996): *Stochastic Programming Problems with Probability and Quantile Functions*, Wiley.

[317] Karatzas, I., Lakner, P., Lehoczky, J.P., and Shreve, S. E., (1991): Equilibrium in a simplified dynamics stochastic economy with heterogeneous agents, in: E. Meyer-Wolf, E. Merzbach, O. Zeitouni (eds), Stochastic analysis Liber Amicorum for Moshe Zakai, Academic Press: New York, 245-272.

[318] Karatzas, I., Lehoczky, J.P., Sethi, S.P., and Shreve, S. E., (1985): Explicit solution of a general consumption/investment problem, *Mathematics of Operations Research*, 25, 157-186.

[319] Karatzas, I., Lehoczky, J.P., Shreve, S.E. and Xu, G.-L., (1991): Martingale and duality methods for utility maximization in an incomplete market, *SIAM Journal of Control and Optimization*, 29, 702-730.

[320] Karatzas, I. and Shreve, S.E., (1991): *Brownian Motion and Stochastic Calculus* (2nd edition), Springer: New York.

[321] Karatzas, I. and Shreve, S.E., (1998): *Methods of Mathematical Finance*, Springer: New York.

[322] Karatzas, I. and Wang, H., (2001): Utility maximization with discretionary stopping, *SIAM Journal of Control and Optimization*, 39, 306-329.

[323] Karni, E., (1985): *Decision Making under Uncertainty: the Case of State-Dependent Preferences*, Harvard University Press: Cambridge, MA.

[324] Kataoka, S., (1963): A stochastic programming model, *Econometrica*, 31, 181-196.

[325] Kazemi, H., Schneeweis, T. and Gupta, R., (2003): Omega as performance measure, CISDM University of Massachusetts, Amherst.

[326] Keating, C. and Shadwick, F., (2002): A universal performance measure, *Journal of Performance Measurement*, 6, 59-84.

[327] Kihlstrom, R. and Mirman, L.J., (1974): Risk aversion with many commodities, *Journal of Economic Theory*, 8, 361-388.

[328] Knight, F.H., (1921): *Risk, Uncertainty and Profit*, Hart, Schaffner & Marx; Houghton Mifflin company: Boston, MA.

[329] Konno, H, Shirakawa, H., and Yamazaki, H., (1993): A mean-absolute deviation-skewness portfolio optimization model, *Annals of Operations Research*, 45, 205-220.

[330] Konno, H. and Wijayanayake, A., (2001): Portfolio optimization problem under concave transaction costs and minimal transaction unit constraints. *Mathematical Programming*, 89, 233-250.

[331] Konno, H. and Yamazaki, H., (1991): Mean-absolute deviation portfolio optimization models and its applications to Tokyo stock market, *Management Science*, 37, 519-531.

[332] Koopmans, T.C., (1960): Stationary ordinal utility and impatience, *Econometrica*, 28, 287-309.

[333] Korn, R., (1997): *Optimal Portfolios - Stochastic Models for Optimal Investment and Risk Management*, World Scientific.

[334] Korn, R., (1998): Portfolio optimization with strictly positive transaction costs and impulse control, *Finance and Stochastics*, 2, 85-114.

[335] Kramkov, D. and Schachermayer, W., (1999): The asymptotic elasticity of utility functions and optimal investment in incomplete markets, *Annals of Applied Probability*, 9, 904-950.

[336] Krokhmal, P., Palmquist, J. and Uryasev, S., (2001): Portfolio optimization with conditional Value-at-Risk objective and constraints, *Journal of Risk*, 4, 43-68.

[337] Krokhmal, P., Uryasev, S., and Zrazhevsky G., (2002): Risk management for hedge fund portfolios: a comparative analysis of linear rebalancing strategies, *Journal of Alternative Investments*, 10-29.

[338] Kroll, Y., Levy, H., and Markowitz, H., (1984): Mean-variance versus direct utility maximization, *Journal of Finance*, 39, 47-61.

[339] Krylov, N.V., (1980): *Controlled Diffusion Process*, Springer: Berlin.

[340] Kushner, H.J., (1967): Optimal discounted stochastic control for diffusion processes, *SIAM Journal of Control*, 5, 520-531.

[341] Kushner, H.J. and Dupuis, P.G., (1992): *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer: New York, Berlin.

[342] Kusuoka, S., (2001): On law invariant coherent risk measures, in: *Advances in Mathematical Economics*, Springer: Tokyo, 3, 83-95.

[343] Lakner, P., (1998): Optimal trading strategy for an investor: the case of partial information, *Stochastic Processes and their Applications*, 76, 77-97.

[344] Larsen, N., Mauser H., and Uryasev, S., (2002): Algorithms for optimization of Value-at-Risk, in: P.Pardalos and V.K. Tsitsiringos, (Eds) Financial Engineering, e-commerce and Supply Chain. Kluwer Academic Publishers, 129-157. (see http://www.ise.ufl.edu/uryasev/VaR-minimization.pdf).

[345] Lederman, J. and Klein, R., (1994): *Global Asset Allocation: Techniques for Optimizing Portfolio Management*, Wiley.

[346] Ledoit, O. and Wolf, M., (2002): Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size, *Annals of Statistics*, 30, 1081-1102.

[347] Leland, H.E., (1980): Who should buy portfolio insurance? *Journal of Finance*, 35, 581-594.

[348] Leland, H.E., (1999): Beyond mean-variance: risk and performance measurement in a nonsymmetrical world, *Financial Analysts Journal*, 54, Jan./Feb.

[349] Leland, H.E. and Rubinstein, M., (1988): The evolution of portfolio insurance, in: D.L. Luskin, ed., Portfolio Insurance: a Guide to Dynamic Hedging, Wiley.

[350] Lehoczky, J.P., Sethi, S.P., and Shreve, S.E., (1983): Optimal consumption and investment policies allowing consumption constraints and bankruptcy, *Mathematics of Operations Research*, 8, 613-636.

[351] Levkovitz, R. and Mitra, G., (1993): Solution of large scale linear programs: a review of hardware, software and algorithmic issues, *Optimization in industry*, Wiley, 139-172.

[352] Lévy, H., (1994) : Absolute and relative risk aversion: An experimental study. *Journal of Risk and Uncertainty*, 8, 289-307.

[353] Lévy, H. and Lévy, M., (2004) : Prospect theory and mean-variance analysis, *Review of Financial Studies*, 17, 1015-1041.

[354] Lhabitant, F.-S., (2001): Hedge funds investing: a quantitative look inside the black box, working paper, Union Bancaire Privée.

[355] Lhabitant, F.-S., (2004): *Hedge Funds: Quantitative Insights*, Wiley.

[356] Lichtenstein, S. and Slovic, P., (1971): Reversal of preferences between bids and choices in gambling decisions, *Journal of Experimental Psychology*, 89, 46-55.

[357] Lioui, A. and Poncet, P., (2001): On optimal portfolio choice under stochastic interest rates, *Journal of Economic Dynamics and Control*, 25, 1841-1865.

[358] Lions, P.-L., (1983): Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations. Part II: Viscosity solutions and uniqueness, Comm. Partial Differential Equations, 8, 1229-1276.

[359] Lo, A., (2002): The statistics of Sharpe ratios, *Financial Analysts Journal*, 58, 36-52.

[360] Lobosco, A., (1999): Style/risk-adjusted performance, *Journal of Portfolio Management*, 25, 65-68.

[361] Long, J.B., (1990): The numeraire portfolio, *Journal of Financial Economics*, 26, 29-69.

[362] Longin, F., (2001): Beyond the VaR, *Journal of Derivatives*, Summer, 36-48.

[363] Longin, F. and Solnik, B., (1995): Is the correlation in international equity returns constant: 1960-1990?, *Journal of International Money and Finance*, 14, 3-26,

[364] Longin, F. and Solnik, B., (2001): Extreme correlation of international equity markets, *Journal of Finance*, 56, 651-678.

[365] Loomes, G. and Sugden, R., (1982): Regret theory: an alternative theory of rational choice under uncertainty, *Economic Journal*, 92, 805-824.

[366] MacFadden, D., (1973): Conditional logit analysis of qualitative choice behavior, in Zarembka, P. (ed.), *Frontiers in Econometrics*, Academic Press: New York.

[367] Machina, M., (1982a): Expected utility analysis without the independence axiom, *Econometrica*, 50, 277-323.

[368] Magill, M.J.P., and Constantinides, G.M., (1976): Portfolio selection with transaction costs, *Journal of Economic Theory*, 13, 264-271.

[369] Maccheroni, F., Marinacci, M. and Rustichini, A., (2005): Dynamic variational preferences, working paper Fondazione Collegio Carlo Alberto.

[370] Macovschi, S. and Quittard-Pinon, F., (2005): Pricing power and other polynomial options, to appear in *Journal of Derivatives*.

[371] Malevergne, Y. and Sornette, D., (2004): High-Order moments and cumulants of multivariate asset returns: analytical theory and empirical tests- II, *Finance Letters*.

[372] Malevergne, Y. and Sornette, D., (2005): Higher-moment portfolio theory: capitalizing on behavioral anomalies of stock markets, *Journal of Portfolio Management*, 31, 49-55.

[373] Markowitz, H.M., (1952a): Portfolio selection, *Journal of Finance*, 7, 77-91.

[374] Markowitz, H.M., (1952b): The utility of wealth, *Journal of Political Economy*, 60, 151-158.

[375] Markowitz, H.M., (1959): *Portfolio Selection: Efficient Diversification of Investments*, Wiley: New York.

[376] Markowitz, H.M., (1987): *Mean-variance Analysis in Portfolio Choice and Capital Markets*, Basil Blackwell: Cambridge, MA.

[377] Marschak, J., (1938): Money and the theory of asset, *Econometrica*, 6, 311-325.

[378] Marschak, J., (1950): Rational behaviour, uncertain prospects and measurable utility, *Econometrica*, 18, 111-141.

[379] Mausser, H. and Rosen, D., (1999): Efficient risk/return frontiers for credit risk, ALGO Research Quarterly, 2, 35-48. (see www.algorithmics.com/research/dec99/arq2-4-frontiers.pdf).

[380] McNeil, A., (1997): Estimating the tails of loss severity distributions using extreme value theory, *ASTIN Bulletin*, 27, 117-137.

[381] McNeil, A. and Frey, R., (2000): Estimation of tail-related risk for heteroscedastic financial time series: an extreme value approaches, *Journal of Empirical Finance*, 7, 271-300.

[382] Meade, N. and Salkin, G.R., (1989): Index funds-construction and performance measurement, *Journal of the Operational Research Society*, 40, 871-879.

[383] Meade, N. and Salkin G.R., (1990): Developing and maintaining an equity index fund,*Journal of the Operational Research Society*, 41, 599-607.

[384] Menger, C., (1871): Grundsätze der volkswirthschaftslehre, Vienna: Braumnalter. Translated as Principles of Economics, by J. Dingwall and B.F. Hoselitz, Glencoe, Ill.: Free Press, 1951. Reprinted in 1976 and 1981, New York University Press: New York.

[385] Merton, R.C., (1969): Lifetime portfolio selection under uncertainty: the continuous time case, *Review of Economics and Statistics*, 51, 247-257.

[386] Merton, R.C., (1971): Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory*, 3, 373-413.

[387] Merton, R.C., (1972): An analytic derivation of the efficient portfolio frontier, *JJournal of Financial Quantitative Analysis*, 7, 1851-1872.

[388] Merton, R.C., (1980): On estimating the expected return on the market, *Journal of Financial Economics*, 8, 323-361.

[389] Merton, R.C., (1990): *Continuous Time Finance*, Basil Blackwell: Cambridge, MA.

[390] Meucci, A., (2005): *Risk and Asset Allocation*, Springer: Berlin.

[391] Michaud, R.O., (1998): *Efficient Asset Management: A Practical Guide to Stock*, *Portfolio Optimization and Asset Allocation*, Harvard Business School Press: Boston, MA.

[392] Modigliani, F., and Pogue, G.A., (1974): An introduction to risk and return - II, *Financial Analysts Journal,* 30, 69-86.

[393] Morgan, J.P., (1996): Risk Metrics-technical document, 4th edition, J.P. Morgan: New York.

[394] Morton, A. J. and Pliska, S., (1995): Optimal portfolio management with fixed transaction costs, *Mathematical Finance*, 5, 337-356.

[395] Mulvey, J. M., Vanderbei, R. J., and Zenios, S. A., (1995): Robust optimization of large scale systems, *Operations Research*, 43, 264-281.

[396] Muralidhar, A.S., (2001): Optimal risk-adjusted portfolios with multiple managers, *Journal of Portfolio Management*, 27, 97-104.

[397] Nagai, H. and Peng, S., (2001): Risk-sensitive dynamic portfolio optimization with partial information on infinite time horizon, *Annals of Applied Probability*, 173-195.

[398] Nelson, D.B., (1991): Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, 59, 347-370.

[399] Nesterov, Y. and Nemirovski, A., (1995): *Interior-Point Polynomial Algorithms in Convex Programming*, Society for Industrial and Applied Mathematics: Philadelphia.

[400] von Neumann, J. and Morgenstern, O., (1944): *Theory of Games and Economic Behavior*, Princeton University Press.

[401] Oksendal, B., (1995): *Stochastic Differential Equations*, Springer: New York, 4th edition.

[402] Park, J.M. and Staum, J.-C., (1998): Performance persistence in the alternative industry, Paradigm capital management, working paper.

[403] Peng, S., (1997): Backward SDE and related g-expectations. In: *Backward Stochastic Differential Equations*, N. El Karoui and L. Mazliak (Eds), Pitman Res. Notes Math. Ser., 364, Longman, Harlow, 141-159.

[404] Pennachi, G., (1991): Identifying the dynamics of real interest rates models and inflation: evidence using survey data, *Review of Financial studies*, 4, 53-86.

[405] Perold, A., (1984): Large-scale portfolio optimization, *Management Science*, 30, 1143-1160.

[406] Perold, A., (1986): Constant portfolio insurance, Harvard Business School, unpublished manuscript.

[407] Perold, A. and Sharpe, W., (1988): Dynamic strategies for asset allocation, *Financial Analysts Journal*, 44, 16-27.

[408] Pham, H., (2002): Smooth solutions of optimal investment models with stochastic volatilities and portfolio constraints, *Appl. Math. Optim.*, 46, 55-78.

[409] Pham, H., (2002): Minimizing shortfall risk and applications to finance and insurance problems, *Annals of Applied Probability*, 12, 143-172.

[410] Pliska, S.R., (1986): A stochastic model of continuous trading: optimal portfolios, *Mathematics of Operations Research*, 11, 371-384.

[411] Pope, P.F. and Yadav, P.K., (1994): Discovering errors in tracking error, *Journal of Portfolio Management*, Winter, 27-32.

[412] Pratt, J.W., (1964): Risk aversion in the small and in the large, *Econometrica*, 32, 122-136.

[413] Prigent, J.-L., (2001): Assurance du portefeuille : analyse et extension de la méthode du coussin, *Banque et Marchés*, 51, 33-39.

[414] Prigent, J.-L., (2003): *Weak Convergence of Financial Markets*, Springer: Berlin.

[415] Prigent, J.-L., (2004): Generalized option based portfolio insurance, working Paper Thema, University of Cergy, France.

[416] Prigent, J.-L. and Tahar, F., (2004): Indexed funds management based on TA algorithm and cointegration, working paper Thema, University of Cergy, France.

[417] Prigent, J.-L. and Tahar, F., (2006): Optimal portfolios with guarantee at maturity: computation and comparison, *International Journal of Business*, 11, 171-185.

[418] Prigent, J.-L. and Toumi, S., (2003): Portfolio management with safety criteria in complete financial markets, *International Journal of Business*, 10, 233-250.

[419] Protter, P., (1990): *Stochastic Integration and Differential Equations*, Springer: Berlin.

[420] Quiggin, J., (1982): A theory of anticipated utility, *Journal of Economic Behaviour and Organization*, 3, 323-343.

[421] Quiggin, J., (1991): Comparative statics for rank-dependent expected utility, *Journal of Risk and Uncertainty*, 4, 339-350.

[422] Rachev, S. and Zhang, Y., (2004): Risk attribution and portfolio performance measurements, *Journal of Applied Functional Analysis*, 4, 373-402.

[423] Revuz, D. and Yor, M., (1991): *Continuous Martingales and Brownian Motion*, Springer.

[424] Richard, S., (1975): Optimal consumption, portfolio and life insurance rules for an uncertain lived individual in a continuous time model, *Journal of Financial Economics*, 1, 187-203.

[425] Riedel, F., (2004): Dynamic coherent risk measures, *Stochastic processes and their applications*, 112, 185-200.

[426] Rockafellar, R.T., (1970): *Convex Analysis*, Princeton Mathematics, 28, Princeton University Press.

[427] Rockafellar, R.T. and Uryasev, S.P., (2000): Optimization of conditional Value-at-Risk, *Journal of Risk*, 2, 21-42.

[428] Rockafellar, R.T. and Uryasev, S.P., (2002): Conditional Value-at-Risk for general loss distributions, *Journal of Banking and Finance*, 26, 1443-1471.

[429] Rogers, C. and Talay, D., (eds) (1997): *Numerical Methods in Finance*, Cambridge University Press, Cambridge.

[430] Roll, R., (1977): A critique of the asset pricing theory's tests, *Journal of Financial Economics*, 4, 129-176.

[431] Roll, R., (1992): A mean-variance analysis of tracking error, *The Journal of Portfolio Management*, 18, 13-23.

[432] Roll, R. and Ross, S.A., (1977) : Comments on qualitative results for investment proportions, *Journal of Financial Economics*, 5, 265-268.

[433] Ross, S., (1978): Mutual fund separation in financial theory: the separation distributions, *Journal of Economic Theory*, 17, 254-286.

[434] Ross, S., (1981): Some stronger measures of risk aversion in the small and in the large with applications, *Econometrica*, 49, 621-639.

[435] Rothschild, M. and Stiglitz, J.E., (1970): Increasing risk: I. A definition, *Journal of Economic Theory*, 2, 225-243.

[436] Rothschild, M. and Stiglitz, J.E., (1971): Increasing risk: II. Its economic consequences, *Journal of Economic Theory*, 3, 66-84.

[437] Roy, A.D., (1952): Safety-first and the holding of assets, *Econometrica*, 20, 431-449.

[438] Rubinstein, M., (1976): The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics*, 7, 407-425.

[439] Ross, S.A., (1976): The arbitrage theory of capital asset pricing, *Journal of Economic Theory*, 13, 341-360.

[440] Rudd, A., (1977): A note on qualitative results for investment proportions, *Journal of Financial Economics*, 5, 259-263.

[441] Rudd, A., (1980): Optimal selection of passive portfolios, *Financial Management*, 9/10,57-66.

[442] Rudd, A. and Rosenberg, B., (1979): Realistic portfolio optimization, in: Elton, E. J., Gruber, M. J. (eds.): Portfolio Theory, 25 Years after, Studies in the Management Sciences 11, North-Holland, Amsterdam, 21-46.

[443] Rudolf, M., Wolter, H.-J., and Zimmermann, H., (1999): A linear model for tracking error minimization, *Journal of Banking and Finance*, 23, 85-103.

[444] Ruszczynski, A. and Vandenbei, R., (2003): Frontiers of stochastically nondominated portfolios, *Econometrica*, 71, 1287-1297.

[445] Samuelson, P.A., (1950): Probability and attempts to measure utility, *Economic Review*, 1, 167-173.

[446] Samuelson, P.A., (1952): Probability, utility and the independence axiom, *Econometrica*, 20, 670-678.

[447] Samuelson, P.A., (1969): Lifetime portfolio selection by dynamic stochastic programming, *The Review of Economics and Statistics*, *51*, 239-246.

[448] Satchell, S.E. and Hwang, S., (2001): Tracking error: ex ante versus ex post measures, *Journal of Asset Management*, 2, 241-246.

[449] Savage, L.J., (1954): *The Foundations of Statistics*, Wiley: New York.

[450] Scaillet, O., (2004): Non-parametric estimation and sensitivity analysis of expected shortfall, *Mathematical Finance*, 14, 115-129.

[451] Schachermayer, W., (2000): Optimal investment in incomplete financial markets, *Mathematical Finance: Bachelier Congress 2000*, H. Geman, D. Madan, St. Pliska, T. Vorst, editors, 427-462.

[452] Schachermayer, W., (2001): Portfolio optimization in incomplete financial markets, *Lecture Notes of "Scuola Normale Superiore di Pisa"*.

[453] Scheikh, A., (1996): Barra's risk model, Barra Research Insights.

[454] Schmeidler, D., (1989): Subjective probability and expected utility without additivity, *Econometrica*, 57, 571-587.

[455] Schoemaker, P.J.H., (1993): Determinants of risk-taking: behavioral and economic views, *Journal of Risk and Uncertainty*, (Historical Archive), 6, 49-73.

[456] Schneeweis, T. and Spurgin, R., (2001): The benefits of index option-based strategies for institutional portfolios, *Journal of Alternative Investments*, 44-52.

[457] Segal, U., (1987): The Ellsberg paradox and risk aversion: An anticipated utility approach, *International Economic Review*, 28, 175-202.

[458] Segal, U., (1989): Anticipated utility: a measure representation approach, *Annals of Operations Research*, 19, 359-373.

[459] Seierstad, A. and Sydsaeter, K., (1987): *Optimal Control Theory with Economic Applications*, North-Holland: Amsterdam.

[460] Sethi, S.P., (1997): *Optimal Consumption and Investment with Bankruptcy*, Kluwer.

[461] Shanken, J., (1992): On the estimation of beta-pricing models, *Review of Financial Studies*, 5, 1-33.

[462] Sharpe, W., (1963): A simplified model for portfolio analysis, *Management Science*, 277-293.

[463] Sharpe, W., (1964): Capital asset prices: a theory of market equilibrium under conditions of risk, *Journal of Finance*, 19, 425-442.

[464] Sharpe, W., (1966): Mutual fund performance, *Journal of Business*, 39, 119-138.

[465] Sharpe, W., (1992): Asset allocation management style and performance measurement, *Journal of Portfolio Management*, 18, 7-19.

[466] Sharpe, W., (1994): The Sharpe ratio, *Journal of Portfolio Management*, 21, 49-58.

[467] Sharpe, W., (1998): Morningstar's risk-adjusted ratings, *Financial Analysts Journal*.

[468] Shirakawa, H., (1994): Optimal consumption and portfolio selection with upper and lower bounds cosntraints, *Mathematical Finance*, 4, 1-24.

[469] Shiryaev, A., (1984): *Probability*, Springer: New York.

[470] Shreve, S.E. and Soner, H.M., (1994): Optimal investment and consumption with transaction costs, *Annals of Applied Probability*, 4, 609-692.

[471] Sims, C.A., (1980): Macroeconomcs and reality, *Econometrica*, 48.

[472] Solnik, B., (1974): The international pricing of risk: an empirical investigation of the world capital market structure, *Journal of Finance*, 365-378.

[473] Sorensen, C., (1999): Dynamic asset allocation and fixed income management, *Journal of Financial and Quantitative Analysis*, 34, 513-531.

[474] Sorenson, E.H., Miller, K.L. and Samak V., (1998): Allocating between active and passive management, *Financial Analysts Journal*, 54, 18-31.

[475] Sortino, F., (2001): From Alpha to Omega, in *Managing Downside Risk in Financial Markets.* Frank A. Sortino and Stephen E. Satchell, eds., Reed Educational and Professional Publishing Ltd., 2001.

[476] Sortino, F. and van der Meer, R.A.H., (1991): Downside risk, *Journal of Portfolio Management*, August, 27-31.

[477] Sortino, F. and Price, L.N., (1994): Performance measurement in a downside risk framework, *Journal of Investing*, 3, 59-64.

[478] Stein, C., (1955): Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *Proceedings of the 3rd Berkeley Symposium on Probability and Statistics*, 197-206.

[479] Steinbach, M.C., (1999): Markowitz revisited: single-period and multi-period mean-variance models, working paper SC-99-30, Konrad-Zuse-Zentrum für Informationstechnik Berlin.

[480] Stevens, G., (1998): On the inverse of the covariance matrix in portfolio analysis, *Journal of Finance*, 53, 1821-1827.

[481] Stock, J.H. and Watson, M.W., (1988): Testing for common trends, *Journal of the American Statistical Association,* 83, 1097-1107.

[482] Stulz, R., (1998): *Derivatives, Risk Management and Financial Engineering*, Southwestern Publishing.

[483] Svensson, L., (1989): Portfolio choice with non-expected utility in continuous-time, *Economic letters*, 30, 313-317.

[484] Szegö, G., (2002): Measures of risk, *Journal of Banking and Finance*, 26, 1253-1272.

[485] Szegö, G., (2004): *Risk Measures for the 21st Century*, Wiley: New York.

[486] Tabata, Y. and Takeda, E., (1995): Bicriteria optimization problem of designing an index fund, *Journal of the Operational Research Society*, 46, 1023-1032.

[487] Tahar, F., (2005): Eléments de gestion de portefeuille: analyse et extension de méthodes en gestion indicielle et garantie, PhD Thesis, University of Cergy, France.

[488] Tallon, J-M., (1997): Risque microéconomique et prix d'actifs dans un modèle d'équilibre général avec espérance d'utilité dépendante du rang, *Finance*, 18, 139-153.

[489] Tasche, D., (2002): Expected shortfall and beyond, *Journal of Banking and Finance*, 26, 1519-1533.

[490] Telser, L., (1955): Safety first and hedging, *Review of Economic and Statistics*, 23, 1-16.

[491] Teo, K.L. and Yang, X.Q., (2001): Portfolio selection problem with minimax type risk function, *Annals of Operations Research*, 101, 333-349.

[492] Toumi, S., (2003): Application de la théorie de l'équilibre général et des notions de Value-at-Risk à la finance de marché, Phd Thesis, THEMA, University of Cergy, France.

[493] Toy, W.M. and Zurach, M.A., (1989): Tracking the Euro-Pac index, *Journal of Portfolio Management*, 15, 55-58.

[494] Treynor, J., (1965): How to rate management of investment funds, *Harvard Business Review*, 43, 63-75.

[495] Treynor, J. and Mazuy, K., (1966): Can mutual funds outguess the market? *Harvard Business Review*, 43, 63-75.

[496] Tversky, A. and Kahneman, D., (1992): Advances in prospect theory: cumulative representation of uncertainty, *Journal of Risk and Uncertainty*, 5, 297-323.

[497] Van Laarhoven, P.J.M. and Aarts, E.H.L., (1987): *Simulated Annealing: Theory and Applications*, Kluwer Academic Publishers.

[498] Vasicek, O., (1977): An equilibrium characterization of the term structure, *Journal of Financial Economics*, 5, 177-88.

[499] Wagner, N., (2002): On a model of portfolio selection with benchmark, *Journal of Asset Management*, 3, 55-65.

[500] Wagner, N. (2003): Methods of relative portfolio optimization, in: Satchell, S. E., Scowcroft, A. (eds.): New Advances in Portfolio Construction and Implementation, Butterworth and Heinemann, Oxford, pp. 333-341.

[501] Wakker, P., (1994): Separating marginal utility and risk aversion, *Theory and Decision*, 36, 1-44.

[502] Wang, T., (1996): A characterization of dynamic risk measures, WP Faculty of Commerce and Business Administration, U.B.C.

[503] Weber, S., (2005): Distribution-invariant risk measures, information, and dynamic consistency, *Mathematical Finance*, 16, 419-44.

[504] Wilson, T.C., (1996): Calculating risk capital. In: Alexander, C. (Ed.), The Handbook of Risk Management and Analysis; Wiley, Chichester, 193-232.

[505] Winker, P., (2000): *Optimization Heuristics in Econometrics: Applications of Threshold Accepting,* Wiley, New York.

[506] Yaari, M.E., (1969): Some measures of risk aversion and their uses, *Journal of Economic Theory*, 1, 315-329.

[507] Yaari, M.E., (1987): The dual theory of choice under risk. *Econometrica*, 55, 95-115.

[508] Young, M.R., (1998): A minimax portfolio selection rule with linear programming solution, *Management Science*, 44, 673-683.

[509] Zakoian, J.-M., (1994): Threshold heteroskedastic models, *Journal of Economic Dynamics and Control*, 18, 931-955.

[510] Zariphopoulou, T., (1992): Investment-consumption models with transaction fees and Markov-chain parameters, *SIAM Journal of Control and Optimization*, 30, 613-636.

[511] Zeng, B. and Zhang, J., (2001): An empirical assesement of asset correlation models, KMV LLC, San Francisco.

[512] Zhigljavsky, A., (1991): *Theory of Random Search*, Kluwer Academic Publishers.

Symbol Description

$\mathbb{R}^d$	The real Euclidian space of dimension d
Ω	The set of random events
$\mathbb{P}$	The historical probability
$\mathcal{F}$	The σ-algebra corresponding to known information
$(\mathcal{F}_t)_t$	The filtration corresponding to increasing information along time t
$\mathbb{E}[X]$	Expectation of the random variable X
$\mathcal{L}$	The set of lotteries
$\rho(X)$	Risk measure of the random variable X
RDEU	Rank Dependent Expected Utility
VaR	Value-at-Risk
CVaR	Conditional Value-at-Risk
ES	Expected shortfall
$\succeq$	Preference relation
$X \succeq_i Y$	The random variable X stochastically dominates Y at order i
$\mathbb{L}^p$	Space of all $\mathcal{F}$-measurable random variables such that $\mathbb{E}_{\mathbb{P}}[X^p]$ is finite
$\mathcal{E}$	Doleans-Dade stochastic exponential
R_P	Return of portfolio P
A' or A^t	Transpose of A
$[X, Y]$	Co-variation of processes X and Y
$[X, X]$	Quadratic variation of process X
$\langle X, X \rangle$	Predictable compensator of process X
$\mathbb{I}$	The vector with all components equal to 1
$\mathbb{I}_A$	The indicator function of the subset A
i.i.d.	Independent and identically distributed
r.c.l.l	Right continuous and left limited
w.r.t.	With respect to
CPPI	Constant Proportion Portfolio Insurance
OBPI	Option Based Portfolio Insurance
CARA	Constant Absolute Risk Aversion
CRRA	Constant Relative Risk Aversion
HARA	Hyperbolic Absolute Risk Aversion
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
SDE	Stochastic Differential Equation
BSDE	Backward Stochastic Differential Equation

Index